

**Publication No.: 2001-527304**

**Publication date: December 25, 2001**

**Title of Invention: A method of hierarchical digital video summarization  
and browsing**

5      **Applicant: SHARP INC.**

Disclosed is a method of hierarchical digital video summarization and  
browsing. The method comprises the step of: inputting a digital video signal for a  
digital video sequence; and generating a hierarchical keyframe summary. The method  
10 further comprises the steps of: computing histograms for the digital video sequence;  
detecting shot boundaries within the digital video sequence; determining the number  
of keyframes allocated within each shot; locating the actual position of each keyframe  
within each shot; admitting keyframes by the largest consecutive difference criteria;  
pruning keyframes for a shot without meaningful action; extracting keyframes with a  
15 better efficiency in case of a compression video; and browsing a shot using the  
hierarchical keyframe summary.

**Best Available Copy**

(19) 日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2001-527304

(P2001-527304A)

(43) 公表日 平成13年12月25日 (2001. 12. 25)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テマコード* (参考)
H 0 4 N 7/24		G 0 6 F 17/30	1 7 0 D 5 B 0 5 0
G 0 6 F 17/30	1 7 0		2 2 0 A 5 B 0 7 5
	2 2 0	G 0 6 T 1/00	2 0 0 A 5 C 0 5 9
G 0 6 T 1/00	2 0 0	H 0 4 N 7/13	Z
審査請求 有 予備審査請求 有 (全 58 頁)			

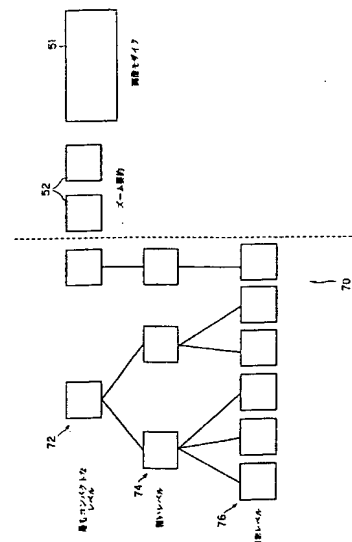
(21) 出願番号 特願2000-525829(P2000-525829)  
 (86) (22) 出願日 平成10年12月17日(1998. 12. 17)  
 (85) 翻訳文提出日 平成12年4月28日(2000. 4. 28)  
 (86) 国際出願番号 P C T / J P 9 8 / 0 5 7 1 5  
 (87) 国際公開番号 W O 9 9 / 3 2 9 9 3  
 (87) 国際公開日 平成11年7月1日(1999. 7. 1)  
 (31) 優先権主張番号 0 8 / 9 9 4 , 5 5 8  
 (32) 優先日 平成9年12月19日(1997. 12. 19)  
 (33) 優先権主張国 米国 (U S)  
 (81) 指定国 E P (A T , B E , C H , C Y ,  
 D E , D K , E S , F I , F R , G B , G R , I E , I  
 T , L U , M C , N L , P T , S E ) , J P , K R

(71) 出願人 シャープ株式会社  
 大阪府大阪市阿倍野区長池町22番22号  
 (72) 発明者 ラタコンダ, クリッシュナ  
 アメリカ合衆国, 61801 イリノイ, アバ  
 ーナ, アパートメント, 6, 1010 ダブリ  
 ュ, クラーク  
 (74) 代理人 弁理士 高野 明近 (外2名)  
 Fターム(参考) 5B050 AA09 BA06 DA01 EA10 EA18  
 GA08  
 5B075 ND12 NK08 NR12 NR16 NS01  
 PQ02  
 5C059 KK36 KK39 MA00 MC11 PP04  
 SS12 TA80 TB02 TC11 TC14  
 TD10 UA05

(54) 【発明の名称】 デジタル動画の階層的要約及び閲覧方法

## (57) 【要約】

デジタル動画の階層的要約及び閲覧方法は、デジタル動画シーケンス用のデジタル動画信号の入力と動画シーケンスのキーフレームに基づく階層的要約を生成することを含んでいる。追加のステップとして、デジタル動画シーケンスに対するヒストグラムを計算することと、デジタル動画シーケンス内のショット境界を検出することと、各ショット内に配分されるキーフレーム数を算定することと、各ショット内の各キーフレームの動きのある位置を決定することと、最大連続差分判定基準によりキーフレームを認定することと、有意な作の無いショット毎にキーフレームを削減することと、圧縮動画の場合にキーフレームを効率よく抽出することと、階層的キーフレームの要約を用いてショットを閲覧することを含んでいる。



## 【特許請求の範囲】

【請求項1】 デジタル動画シーケンスに対するデジタル動画信号を入力し、階層的キーフレームの要約を生成することを含んでなる階層的デジタル動画要約及び閲覧方法。

【請求項2】 前記入力した後に、前記デジタル動画シーケンスに対するヒストグラムを計算し、前記デジタル動画シーケンス内のショット境界を検出し、各ショット内に割り当てられたキーフレーム数を決定し、最大連続差分判定基準による各ショット内の各キーフレームの動きのある局面を位置決めし、有意な動きの無いショットに対するキーフレームを削減することを含むことを特徴とする請求項1に記載の階層的デジタル動画要約及び閲覧方法。

【請求項3】 前記生成した後に、前記階層的キーフレームの要約を用いるキーフレームを閲覧することを含むことを特徴とする請求項2に記載の階層的デジタル動画要約及び閲覧方法。

【請求項4】 前記入力した後に、ディゾルブイベントを検出し除去することを含むことを特徴とする請求項2に記載の階層的デジタル動画要約及び閲覧方法。

【請求項5】 前記ディゾルブイベントを検出し除去した後に、パンイベントとズームイベントよりなるイベント群より取ったイベントを含む前記デジタル動画シーケンス内のフレームを検出することによって、グローバルな動きを検出することを含むことを特徴とする請求項4に記載の階層的デジタル動画要約及び閲覧方法。

【請求項6】 パンイベントを検出し、画像モザイクを形成することを含むことを特徴とする請求項5に記載の階層的デジタル動画要約及び閲覧方法。

【請求項7】 ズームイベントを検出し、前記イベント中のズームの度合を推定し、ズーム要約をコンパイルすることを含むことを特徴とする請求項5に記載の階層的デジタル動画要約及び閲覧方法。

【請求項8】 前記階層的要約のプロセスからグローバルな動きイベントを除外することを含むことを特徴とする請求項5に記載の階層的デジタル動画要約及び閲覧方法。

【請求項9】 前記階層的キーフレームの要約を生成することは、前記階層的キーフレームの要約を、最もコンパクトなレベルの要約と粗いレベルの要約と最も細密なレベルの要約とを含む複数レベルの要約に分割することを含むことを特徴とする請求項1に記載の階層的デジタル動画要約及び閲覧方法。

【請求項10】 特定レベルの要約を選択した後に、ユーザによって前記キーフレームを閲覧することを含むことを特徴とする請求項9に記載の階層的デジタル動画要約及び閲覧方法。

【請求項11】 前記キーフレームの階層的要約中のキーフレームを、蓄積、検索或いは表示のために、サムネイルに空間的にサブサンプリングできることを特徴とする請求項9に記載の階層的デジタル動画要約及び閲覧方法。

【請求項12】 前記階層的キーフレーム要約を生成することは、キーフレームをクラスタリングし、より粗いレベルの要約のキーフレームを生成することを含むことを特徴とする請求項9に記載の階層的デジタル動画要約及び閲覧方法。

【請求項13】 前記クラスタリングは、より粗いレベルにおけるキーフレーム数のコンパクション率を生成することを含むことを特徴とする請求項12に記載の階層的デジタル動画要約及び閲覧方法。

【請求項14】 前記クラスタリングは、対ごとのクラスタリングを含むことを特徴とする請求項12に記載の階層的デジタル動画要約及び閲覧方法。

【請求項15】 前記のより粗いレベルの要約のキーフレームを生成することとは、最大連続差分判定基準を用いてキーフレームを生成することを含むことを特徴とする請求項12に記載の階層的デジタル動画要約及び閲覧方法。

【請求項16】  $t_0 = 0$ 、及び第2フレームを第1キーフレームとなるための候補として選択するとして、 $k_1 = 1$ を設定し、 $i = 1 \sim K-1$ に対して、 $t_i = 2k_i - t_{i-1}$ である $k_{i+1}$ を $2C(t_i) - C(k_i) \leq C(k_{i+1})$ が維持される第1動画フレームとして定義し、 $i = K$ に対して、 $t'_K = 2k_K - t_{K-1}$ を計算し、 $t_K > 2k_K - t_{K-1} = t'_K$ でなければ、以前の反復結果を維持して全ての $k_i$ に $t_K = t'_K$ になるようにオフセット値を与えて終了し、それ以外であれば、 $k_i$ を1だけ増加させて前記定義する処理に進むことにより、キーフレームを

認定することを含むことを特徴とする請求項9に記載の階層的デジタル動画要約及び閲覧方法。

【請求項17】 前記計算することは、前記ショットの最後のキーフレームを  $t_{k-1}$  と  $t_k$  の間の中点付近に位置決めすることを含むことを特徴とする請求項16に記載の階層的デジタル動画要約及び閲覧方法。

【請求項18】 前記認定することは、 $(n/K)$  番目毎のフレームをキーフレームとして選択することを含むことを特徴とする請求項16に記載の階層的デジタル動画要約及び閲覧方法。

【請求項19】 前記認定することは、興味を引かないショットを検出し、前記階層的キーフレーム要約から前記ショットのキーフレームを除去することを特徴とする請求項16に記載の階層的デジタル動画要約及び閲覧方法。

【請求項20】 前記入力することは、圧縮デジタル動画シーケンスを入力し、ビットストリームのインデックステーブルを生成することを含み、前記ヒストグラムを計算することは、前記圧縮デジタル動画シーケンスを部分的にのみ復号することを含むことを特徴とする請求項1に記載の階層的デジタル動画要約及び閲覧方法。

【請求項21】 前記各ショット内のキーフレームを割り当てることは、前記キーフレームを完全に復号することを含むことを特徴とする請求項20に記載の階層的デジタル動画要約及び閲覧方法。

【請求項22】 前記キーフレームを完全に復号することは、前記動画ビットストリームを構文解析せず、かつ、ビットストリームインデックステーブルを用いて前記動画ビットストリームを完全に復号することなく前記キーフレームを復号することを含むことを特徴とする請求項21に記載の階層的デジタル動画要約及び閲覧方法。

【請求項23】 DCTに基づいた圧縮動画を前記部分的に復号することは、ヒストグラムを計算するためのDCT係数のDC値を使用することを含むことを特徴とする請求項20に記載の階層的デジタル動画要約及び閲覧方法。

【請求項24】 前記部分的に復号することは、キーフレームと該キーフレームの参照フレームのみを復号することを含むことを特徴とする請求項20に記載

の階層的デジタル動画要約及び閲覧方法。

【請求項25】 前記復号することは、復号器マネージャによって復号することを含むことを特徴とする請求項20に記載の階層的デジタル動画要約及び閲覧方法。

【請求項26】 前記復号器マネージャによって復号することは、ビットストリームインデックステーブルを用いて、全動画ビットストリームの最少の復号及び構文解析でキーフレームを復号することを含むことを特徴とする請求項25に記載の階層的デジタル動画要約及び閲覧方法。

【請求項27】 前記復号器マネージャによって復号することは、削減されたビットストリームインデックステーブルを生成して、前記全動画ビットストリームを構文解析して復号することなくキーフレームを復号するために必要な情報のみを蓄積することを含むことを特徴とする請求項26に記載の階層的デジタル動画要約及び閲覧方法。

## 【発明の詳細な説明】

【0001】

関連出願

1998年1月7日提出のRatakondaの第09/004,058号の“動画シーケンスにおけるディゾルブ領域の検出 (Detecting Dissolve Regions in Video Sequences)”。

【0002】

発明の属する技術分野

本発明は、(i) 動画コンテンツの視覚的な認定、(ii) 動画のインデクシング、(iii) 動画の閲覧ブラウジング及び(iv) 動画の編集等の用途に用いる階層的要約において静止画の各セット毎にデジタル動画シーケンスの表現を決定することに関する。デジタル動画シーケンスは、動画像圧縮標準(MPEG)により圧縮され、その表現は圧縮ビットストリームの最少復号化によって決定される。

【0003】

発明の背景

動画のコンパクトな表現は、多数の情報問い合わせと検索の用途にとって肝要である。かような用途の例には、マルチメディアデータベースのアクセスから動画クリップを介したスキミング(高速前方走査)にまで及ぶ。大多数のこれまでの方法は、所与の動画セグメントを“ショット”に分割することに主として重点が置かれていた。各ショットは、そのショットを要約する1枚のキーフレームにより代表される。かように、全動画を通して閲覧する代わりにこれらの代表的なフレームを見ることができる。ショットの検出は高い精度(>90%)と僅かな脱漏率(<5%)で達成される。ヒストグラムに基づく方法は最も成功率が高く、且つ計算量が最少ですむショット検出方法の1つである。種々のショット検出方法間の比較を文献に見ることができる。これらのスキームの多くは、動画ショット境界を定める際にパン、ズーム、ディゾルブ及びフェード等の幾つかの特殊な状況を考慮している。

【0004】

公知の技法は、ショット境界またはシーン転換を検出し、動画シーケンスを代表するキーフレームのような各ショットからの単一フレームよりなるフレームの集積を使用することに一般的には重点を置いている。各ショットに1枚より多いキーフレームを割り当てることにより、その動画のコンテンツをより良く要約表現できる。かような公知の要約方法は、しかしながら、単層の要約であり柔軟性を欠く。

他の公知の技法は、色ヒストグラムを使用し、MPEGビットストリーム（即ち、DCT 8x8ブロックのDC係数のヒストグラム）からヒストグラムを形成する方法を記述している。これは、I（フレーム内符号化）フレームに対する比較的直截な方法であるが、参照画像の最少復号化を伴うP（予測）フレーム又はB（双方向予測）フレームのDC（ゼロ頻度）係数の回復方法が複数存在する。

離散余弦変換（DCT）圧縮動画に関する公知の文献は、ワーキングシステムの具体的な実施例については何も記述していない。例えば、認定後にキーフレームは視覚表示のために復号されねばならないが、公知の文献のどれにも、全動画シーケンスを復号することなくビットストリームの任意位置にあり得るキーフレームの効率的な復号機構は記述されていない。

#### 【0005】

上記スキームの主要な制限は、全てのショットを同等に扱っていることである。大抵の場合、完全なショットを1枚のフレームでは十分に表現できないと考えられる。これは、ショット中の“興味深い動き”量に応じて、各ショット毎に若干のキーフレームを割り当てるアイデアに到達する。最新技術の動画閲覧システムは、動画シーケンスを成分ショットに分割し、各ショットを数枚の代表キーフレームで表現する。この表現を“要約”と称する。

#### 【0006】

本発明は、1996年のローザンヌ国際画像処理会議の会報（671-674頁）にL. Langendijk, A. Hanjalic, M. Ceccarelli, M. Soletic及びE. Persoonによって開示された“SMASHシステムにおける視覚的探索（Visual Search in SMASH System）”方法（これより後、“Langerdijk”の技



法と記す)を改良し拡張するものである。

【0007】

#### 発明の要約

本発明は、階層的デジタル動画要約及び閲覧方法であり、その基本構成には、デジタル動画シーケンス毎にデジタル動画信号を入力することと、動画シーケンスのキーフレームに基づいて階層的要約を生成することを含んでいる。追加のステップとして、デジタル動画シーケンスに対するヒストグラムを計算することと、デジタル動画シーケンス内のショット境界を検出することと、各ショット内の割り当てられるキーフレーム数を算定することと、各ショット内の各キーフレームの動きのある位置を決定することと、最大連続差分判定基準によりキーフレームを認定することと、有意な動きの無いショット毎にキーフレームを削減することと、圧縮動画の場合にキーフレームを効率よく抽出することと階層的なキーフレームの要約を用いてショットを閲覧することを含んでいる。

【0008】

“動画要約”はその動画の代表として使用できる所与の動画シーケンスの最も突出した(顕著な、主要な)フレームを決定することに関する。階層的要約方法は、レベル(即ち、フレーム数)が細部において異なる、複数のレベルを持つ階層的要約を構築するために開示される。最も粗いか最もコンパクトなレベルは、最も顕著なフレームを提供し、最少数のフレームを含んでいる。

【0009】

本発明の目的は、各レベルが細部の異なるレベルに相当する階層的マルチレベルの要約を生成する方法を提供することにある。

【0010】

本発明の他の目的は、キーフレームの選択を改善する方法を提供することにある。

【0011】

本発明の他の目的は、階層的フレームの要約と共に、シーンの動きコンテンツ特にパンとズームを検出して利用しユーザに提示することにある。

【0012】

本発明のさらなる目的は、各レベルが細部の異なるレベルに相当するMPEG-2による圧縮動画の階層的マルチレベルの要約を生成する方法を提供することにある。

【0013】

本発明のさらなる目的は、ヒストグラムの構築とビットストリームの復号が最少である階層的要約を生成するために、MPEG-2の圧縮動画に直接応用できる方法を提供することにある。

【0014】

本発明の他の目的は、MPEG-2の圧縮動画の要約を生成する完全で効率のよいシステムを提供することにある。

【0015】

本発明のさらに他の目的は、MPEGビットストリームに対するヒストグラム計算を処理する効率のよい方法を提供することにある。

【0016】

好適な実施形態の詳細な説明

セマンテックな顕著さ（特徴）を自動捕捉する現在の技術はまだ未熟であるので、動画の要約方法は、色ヒストグラムのような低レベルな画像の特徴に依拠している。動画の要約は、その動画の代表として使用できる所与の動画シーケンスの最も顕著なフレームを決定する方法である。重要な情報を有する特定のフレームは、総数が予め規定されたフレームを含む単一の要約には含まれないようにすることも可能である。

【0017】

図1を参照する。本発明の階層的要約方法により生成された階層的マルチレベルな要約20は、重要なコンテンツの情報を失わないように十分に多数のフレームを用いた詳細にわたる細密レベルの要約を提供できるので重要なコンテンツ情報は失われない、と同時に、動画の迅速な閲覧と特定用により粗い又はコンパクトな要約の利用を妨げないようにより粗いレベルの詳細度の低い要約を提供できる。階層的マルチレベル要約20は、より詳細な要約が必要と判断されて、粗い要約24及び最も細密な要約26のようなより細かいレベルの要約が呼び出され

るまで提示される、最も粗いレベルの最もコンパクトな要約22を含んでいる。

図1には3つの要約レベルが示されているが、本発明の階層的要約は、1つより多い任意のレベル数を用いることができる。

#### 【0018】

要約20は、又、動画シーケンスのデータベースの高速閲覧を容易にし、最もコンパクトな要約に基づく閲覧ができ、ユーザの要求に従って、より詳細なレベルに要約を高めていくことができる。

#### 【0019】

階層的マルチレベルな要約は、動画の要約を視覚的な対話式で効率よくユーザに提示する方法を容易にする。ユーザはグラフィカルユーザインターフェースを介して要約と対話し、より詳細な要約を要求し、異なるレベルの要約を視覚表示し、任意レベルにおける要約の任意の2枚のキーフレーム間で動画を再生することができる。ここに開示する方法のユーザは要約中のキーフレームの最大数と階層のレベル数を指定できるので、このシステムメモリと資源が制限された用途においても制御可能である。

#### 【0020】

ここに開示する方法は、非圧縮（又は伸長）動画又はMPEG圧縮動画のようなDCT（離散余弦変換）による圧縮動画又は他の動き補償予測圧縮動画のいずれにも適用できる。MPEG圧縮動画の場合、要約は最少のビットストリームの復号と効率的なキーフレームの復号方法により実行され、計算及びメモリ能力に対する要求が軽減される。ここでは、MPEG-2による圧縮動画を例にとり説明するが、前記の通り、DCTに基づく任意の圧縮動画に対しても適用可能である。当技術に熟達した人には、MPEG動画への参照が他に規定されていない限り圧縮動画ストリームへの参照であることは理解されよう。

#### 【0021】

ここに開示の階層的要約方法において、後処理によるフェードのような特殊効果の検出が支援される。かような効果を含むセグメントは、精度に悪影響を与えないように要約プロセスには含まれない。本方法において、要約における最もコンパクトで意味深い表現に対して、パン及びズームセグメントを検出することも

考慮される。

動画シーケンスは、静止画像に対して開発された技法を用いて要約フレームに基づいてインデクシングすることができる。複数レベルにより細部変化レベルにおけるインデクシング（索引）の方向に柔軟性を生じる。

#### 【0022】

本発明の階層的な方法は、最もコンパクトな要約のコンテンツが実際に興味を引くものであればより詳細な要約24、26にアクセスする選択肢を持ちつつ、ユーザが最もコンパクトな要約22を考慮することにより、動画シーケンスの集まりを高速閲覧できる。本発明の方法のユーザは、動画シーケンスの選択セグメントにおいて要約を精錬する柔軟性を有している。

MPEG動画シーケンスを要約するために使用する、“ビットストリームインデックステーブル生成器”と“復号器マネージャ”と云う2つの構成要素が設けられている。これらの構成要素は、視覚的な要約を生成し、次に動画の全体を復号せずに閲覧するようにキーフレームを効率よく復号するために必要である。

#### 【0023】

図2に、本発明の方法全体30を要約して示す。この方法は、カメラ一体型レコーダのような動画カメラとレコーダ又は少なくとも動画シーケンスを設置でき、理想的には動画入力32として使用する大量の動画データを蓄積できる容量を有するコンピュータシステム上で作動することを目的としている。本発明の方法を含む機構をここでは“システム”と呼ぶ。入力動画32は、フェードイン又はフェードアウトのような特殊効果を含むフレームは、擬似ショット境界と擬似キーフレームを生じる結果を招くので、まず、かようなフレームを検出し除去するための処理をする(34)。かようなフレームは大域的(グローバル)な動きイベントとして分類して、次にさらなる処理から除外する。次のステップは、ヒストグラムの計算である(36)。画像色ヒストグラム、即ち、色分布は、動画フレームを代表する特徴ベクトルを構成して、ショット境界の検出38とキーフレームの選択に使用される。ショット境界の検出38は、スレッショルド(閾値)方法を用いて行い、連続するフレームのヒストグラム間の差を比較する。所与のキーフレーム総数(ユーザ指定)40で、各ショットに、良く知られた技法によ

り、ショット内の“動き（アクション）”に応じてキーフレーム数（42）を割り当てる。細密レベルのキーフレームの選択44は、Legendikの技法を改良バージョンを用いて行う。ここに開示する実装は、この方法の改良バージョン、即ち、細密レベルのキーフレームの選択方法を拡張した、後でより詳細に説明する図4に示すような新しい追加のステップを含んだ技法を用いる。

#### 【0024】

図1を再び参照する。画像のモザイク51を生成する結果をもたらす自動パン／ズーム46とズーム要約52は任意選択ステップであり後で説明する。次のステップは、キーフレームの階層化48、即ち、細密レベルの要約より粗い細部での要約を生成する新しい方法である。このプロセスについても後で詳しく説明する。これは、ベクトル量子化の主題の変形に基づいている。細密レベルと粗いレベルの要約は、所与のショットに対し決定され、動画シーケンス中の他のショットに対してもこのプロセスが繰り返される（ブロック50）。追加のステップは、階層的要約の閲覧（ブロック53）とプロセスの終了（ブロック54）を含んでいる。

#### 【0025】

##### 自動パン／ズーム処理

自動パン／ズーム処理ステップ46の詳細は図3に示す通りであり、デジタル動画シーケンスにおけるパン及びズームのイベント（事象）の検出を含んでいる。グローバルな動きを含むフレームを検出する（56）。これは、グローバルな動きの影響を受けるフレームを特定することを目的とした予備的選抜方法である。これらのフレームは、グローバルな動きがパン検出器で検出（58）されるようなカメラのパンによる動きであれば画像モザイクを用いてコンパクトに表示でき、或いは、ズーム要約52、即ち、ズーム検出器60で検出されてズーム評価器（ズーム推定器）66により編集されるような動きであれば、ズームイン又はズームアウトシーケンスの最初と最後のフレームを用いてコンパクトに表示できる。従って、モザイクの形成62は、グローバルなパンの動きを示すこれらのフレームを対象にするだけであり、画像モザイク51をもたらす。画像モザイク51又はズームイン又はズームアウトに関係するフレームは、細密レベルの要約

からは除外される（ブロック64）。理由は、細密レベルの要約はさらに処理してより粗い、よりコンパクトなレベルの要約を形成するからである。

#### 【0026】

他の実施形態において、パン／ズーム処理46は、自動ではなく対話形式で実施できる。ユーザは、選択された細密キーフレーム要約44から、キーフレームを選択できる。それらキーフレームは、パンシーケンスの開始と終了を形成し、システムはこれに応え画像モザイク51を形成し、ユーザに提示する。ユーザは、フレーム番号K及びL、即ち、カメラがパンした始点と終点の2つのキーフレームを認定又はタグ付けすることができる。モザイク形成器62はモザイク形成時にフレーム番号K-nとL+n（nは予め定められたオフセット値）の間のフレームを考慮する。モザイク形成器62は、通常の技術を有する者には良く知られている画像ステッチング技法により実装できる。ズームの場合、パンの場合と同様に、ユーザは開始と終了のフレームを手動で指定できる。自動ズーム検出アルゴリズムを採用してもよいし、このアルゴリズムもまた通常の技術を有する者には良く知られている。

#### 【0027】

図5に階層的要約の形式70を示す。この階層的要約は、階層的キーフレームレベルに分割される。ユーザには先ず最もコンパクトな（最も粗い）レベルの要約72、即ち、最もコンパクトな要約が可能な場合、画像モザイク51及びズーム要約52と共に提示される。次に、ユーザは、ここでは粗いレベル74と称するより細かなレベルで親フレームにタグを付け、子フレームを見ることができる。最も細密なレベル76のフレームのタグ付けにより、動画の再生、例えば、j番目のキーフレームを最も細密なレベルでタグ付けすると、j番目と(j+1)番目の間のフレームが再生される。実際のGUI実装において、親子関係は、表示時に明示できる。ここで用いるように、“タグ”又は“タグ付け”は、特定のフレーム上でクリックするように、コンピュータモニタ上で特定のオブジェクトを認定することにより達成される。階層的要約におけるキーフレームは、蓄積コストを有利にし、要約の高速検索・表示を実現するために、空間的にさらに“サムネール”にサブサンプリングする。動画シーケンスの通常の再生は、最も細密

なレベルで行われるが、より粗いレベルでも実施できる。

【0028】

#### 非圧縮動画の入力

本発明の第1の実施形態は、階層的デジタル動画の要約と閲覧に対する“画素ドメイン”法として記述し、非圧縮動画の入力(32)を要する。

全動画シーケンス用に用いられるキーフレームの総数は所与である(通常蓄積空間要求条件により決まる)と仮定する。Langendikの技法は3つの主要ステップ、即ち、

- 1) ショット境界の検出と、
- 2) 各ショットに割り当てられるキーフレーム数の決定と、
- 3) 各ショット内のキーフレームの位置の発見のステップを持つ。

【0029】

図4に示す本発明に用いられるこの技法80は、3ステップの反復方法(82)を含んでいる。ショット境界の検出(図2の38)は、動的閾値によるヒストグラムに基づく方法を用いてなされる。シーケンスの最初のnフレーム(通常n=3)は、ショット境界に対応しないものと仮定する。平均動き測度 $A_m$ と動き測度の標準偏差 $A_{sd}$ は、最初のnフレームに対する後述する動き測度の平均と標準偏差を各々計算して定める。閾値は $A_m + \alpha A_{sd}$ に設定する。この閾値により境界が検出されれば、次のショット用の新しい閾値を、この新しいショットの最初のnフレームを用いて、同じやり方で決定する。パラメータ $\alpha$ は、通常10に設定する。

2つのヒストグラム( $h_1$ と $h_2$ )間の動き測度( $A(\cdot, \cdot)$ )を(linorm)とする。

【0030】

【式1】

$$A(h_1, h_2) = \sum |h_1(i) - h_2(i)| \quad (1)$$

【0031】

nフレーム $S_1, \dots, S_n$ を持つ1ショットSに対する累積動き測度( $C(\cdot)$ )

)は次式で計算される。

【0032】

【式2】

$$C(x) = \sum_{s_i=s_1}^s A(h_s, h_{s_{i-1}}), (x \leq n) \quad (2)$$

【0033】

各ショット毎の累積動き測度と各ショットの累積動き測度の合計がcaようにして算定される。特定のショット“s”に割り当てられたキーフレーム数（ブロック42）は、そのショット内の累積動き測度の相対量に比例している。

ショット内のキーフレームの動きのある位置の位置決めは1の最小化問題として記述できる。各キーフレームは連続した1セットの動画フレームを代表する（代替する）。動画フレームのこの連続のセットの統合が完全なショットである。連続動画フレームの各セットが1枚のキーフレームで代表されるので、動画フレームの1連続セット内の動き量を小さくしたいと考える。動画フレームの1連続セット内の“動き量”が多すぎると、1枚のキーフレームではその動き量を十分に代表できないことがこの背後にある理由である。caように、1ショットに割り当てられるキーフレームの総数（そのショットが分割された連続セット数と同じ数）が与えられれば、動画フレームの該当連続セット内の“動き”を最少にするキーフレームを発見する最小化手順を使用する。K枚のキーフレームを1ショットs内に位置決めするものと仮定し、キーフレームの位置を $k_j$  ( $j = 1, \dots, K$ )とする。さらに、 $k_j$ のキーフレームで代表される動画フレームの連続セットを $t_{j-1}, \dots, t_j - 1$ とする。換言すると、 $[t_{j-1}, t_j - 1]$ は、キーフレーム $k_j$ により代表されるショットセグメントである。以下のコスト判定条件式が全ての可能な $t_j$ に対して最少となるべきである（ $k_j$ は、 $t_j$ を選択することにより即ち、 $k_j = (t_j + t_{j-1}) / 2$ で決定する）。

【0034】

【式3】

$$g(k_1, \dots, k_K, t_1, \dots, t_{K-1}) = \sum_{j=1}^{j=K} \int_{t_{j-1}}^{t_j} |C(x) - C(k_j)| dx \quad (3)$$



## 【0035】

$t_0$ と $t_K$ はショットの最初と最後のフレームである（従って、定数である）。  
上記コスト判定条件式中に $k_j = (t_j + t_{j-1}) / 2$ を代入し、最適解は、 $2C(t_i) - C(K_i) \leq C(K_{i+1})$ を満足する。

最小化を実行するために、Legendijkの技法と異なる、下記のステップを対話式で実行する。

1.  $k_1 = 1$ に設定する。（ $t_0 = 0$ 、第2フレームを第1キーフレームとなるための候補として選択するものと仮定する。）

2.  $i = 1 \sim K-1$ に対して、

$$t_i = 2k_i - t_{i-1}$$

$k_{i+1}$ を、 $2C(t_i) - C(k_i) \leq C(k_{i+1})$ を満たす第1動画フレーム（即ち、 $t_i$ より大きい最少の下付き添字 $n$ を持つ動画フレーム）と定義する。

3.  $i = K$ に対して、 $t'_K = 2k_K - t_{K-1}$ を計算する。 $t_K > 2k_K - t_{K-1} = t'_K$ であれば、 $k_i$ を1だけ増し、ステップ2に進む。そうでなければ、先の反復結果を保持し、全ての $k_i$ に $t_K = t'_K$ となるようにオフセット値を与えて停止する。

## 【0036】

この最小化は図6に示すように有限数のステップで実行する。累積誤差は1ショット内の非減少関数である。かように、上記の最小化手順は、累積誤差曲線90に対し最良（ $l_i$ に関し最良）の段階近似を与えるキーフレーム $k_j$ を発見することを目的とする。これにより、ショット内の“動き”量に適応変化するキーフレーム $k_j$ の分布が得られる。式3で表現される最小化領域は92である。

## 【0037】

上記の3つのステップの意味は下記の通りである。ショットの最後のキーフレームは $t_{K-1}$ と $t = t_K$ の間の中点にできるだけ近づけなければならない。 $k_i$ を増加し、最初はこの中点を超えるまでステップ2と3を反復し、その後先の反復結果をとり、最後のキーフレームが中点と一致する、即ち、3ステップ反復法により決定される $t_K = 2k_K - t_{K-1}$ と $t'_K$ が $t_K$ と一致するようにオフセット値を与える。

## 【0038】

先のアルゴリズムに導入したもう1つの新規点は、特定のショットに十分多数のキーフレームを割り当てたことにより  $k_i = 1$  であってもショット境界を超えられる場合に関する。この場合、簡単なスキームを用いてキーフレームが等間隔になるように配分する。この簡単なスキームにおいて、1ショットが  $n$  枚のフレームを有し、 $K$  枚のフレームを割り当てる場合、 $(n/K)$  枚毎のフレームをキーフレームとして選択する。

## 【0039】

キーフレームの選択に関する改善

L a g e n d i j k 技法において、 $t_{j-1}$  と  $t_j - 1$  が与えられた1ショットセグメント  $[t_{j-1}, t_j - 1]$  に対するキーフレームは常に  $k_j = (t_j + t_{j-1}) / 2$  に位置する。言い換えれば、キーフレームは、常に、そのセグメントのフレームの代表としてセグメントの中央にあるように選択される。しかしながら、累積誤差の定義に戻れば、累積誤差は連続するフレーム間の絶対変化にのみ依存する。かように、1セグメントの中央の1枚のキーフレームは1フレームより多いフレームに分離された2枚のフレーム間の実際の変化を現わすことはできない。1名のレポートが話している動画シーケンスを考える。例えば10フレーム離れた2枚のフレームがあり、両フレームは、口を開いているレポートを示していると仮定する。結果として、2枚のフレームは、非常に小さい変化又は“動き”を表しているのは明らかである。しかしながら、この2枚のフレーム間の累積変化は連続するフレーム間の絶対変化の合計を表すので前記2枚のフレーム間の累積変化は大きい。これら2枚のフレームは、“セグメントの中央”規則が適用される場合、キーフレームとして選択されたフレームの内にあることが可能である。かように、“セグメントの中央”フレームをキーフレームとして闇雲に選択するとキーフレームの選択を誤ることになる。

## 【0040】

この実施形態の場合、3ステップ反復法によって得られた各ショット内の区切り点の結果集合  $\{t_0, t_1, \dots, t_k\}$  を考慮する。先行キーフレーム ( $k_{j-1}$ ) から (動き測度  $A(,)$  の点で) 最も異なるセグメント ( $t_{j-1}, t_j - 1$ ) 中の

フレームを $k_j$ に位置するキーフレームとして選択する。この方法は、先行キーフレームからの最大差を取るの、ここでは“最大連続差分”基準（ブロック84）と記す。第1キーフレーム（ $k_1$ ）は、3ステップ反復法により決定された1フレームとする。この方法は、連続キーフレームが互いに十分異なり、従って、冗長度をできる限り減少させることを保証する。

【0041】

有意な動きのないショットセグメント中のキーフレーム数の低減

L a g e n d i j kの技法は、上述したように累積誤差に完全に依拠しているために、実際には極めて近似の2枚のフレーム間に大きな誤差があると報告することがあり得る。上述の技法は、与えられたショットセグメント中の最も興味を引くフレームを選択するには良いが、ショットセグメント内の動きの観点からショット全体が“興味を引かない”状況は解決しない。例えば、連続キーフレーム間の十分有意な変化をもたらさない僅かなカメラの動きによる誤差の蓄積があり得る。

有意動きの無いショットを無視するために、それらのショットを認定し、それらショット用キーフレームを削減し（ブロック86）、最も細密なレベルのキーフレームを残す（ブロック44）。これは、2枚の所与のキーフレーム間にある連続動画シーケンスフレーム間で測定して分析された動き測度の平均と標準偏差を推定することによりなされる。2枚のキーフレーム間に十分“有意な動き”が存在すれば、原動画シーケンス中の連続フレーム間の動き測度は有意であり、即ち、そのキーフレームを最大連続差分基準（最大連続差分基準）に従い認定する（ブロック84）。

$A_m$ がフレーム $k_i$ 及び $k_{i-1}$ 間の平均動き測度であり、 $A_{sd}$ を同動き測度の標準偏差とすると次式が得られる。

【0042】

【式4】

$$(s/\beta) * A_{sd} + A_m < A(k_i, k_{i-1}) \quad (4)$$

【0043】

ただし、2枚のキーフレーム $k_i$ 及び $k_{i-1}$ 間の動画シーケンスフレーム数を $s$ とし、2枚のキーフレーム間のコンテンツが興味深い場合である。ショットセグメントが前記の観点から興味を引かないものであれば、その特定のキーフレームを除去しそのショットセグメントを次のショットセグメントに結合する。

上式中のパラメータ $\beta$ は定数である。 $\beta$ が1より小さければ、大きな差異をもつキーフレームのみを残し、過度の削減を招く。ここで報告するシミュレーションの場合 $\beta$ の値は2.0に選定する。ショットに割り当てられたキーフレーム数が小さい場合は、キーフレーム間の距離、従って、キーフレーム間のフレーム数 $s$ が増大するために、ショット割り当てキーフレーム数が小さければ数量 $(s/\beta)$ は増大する。 $(s/\beta)$ が到達し得る最大値を、 $\alpha$ に設定する。この $\alpha$ は、キーフレームの削減数を制限するためにショット境界検出のための閾値を規定する際に使用する係数である。

#### 【0044】

さらなる実験により、直線閾値スキームは、キーフレームの総数選択に不等なキーフレームの割り当てをもたらすことが明らかになった。この問題を軽減するために、上記削減法により消去できるキーフレームの総数の最大百分率に制限値 $MAXERASE=0.3$ を設定する。この制限の場合、 $MAXERASE=0.3$ に相当するフレームの最大冗長度30%が除去される。“最大冗長度”の意味は、 $A(k_i, k_{i-1})$ が最も冗長なキーフレームに対して最少であると云うことを意味する。この場合、 $k_i$ は除去される冗長キーフレームである。

この技法は、高動きシーケンスに応用すると、動きは殆ど構成的であり、即ち、全ての消去されるキーフレームに対して式(4)は満足されており、従って、冗長度は皆無である。

#### 【0045】

#### 階層的要約と閲覧

前記の開示は、知的な“動画インデクシング”システムを記述しているが、かようなシステムは、動画フレームの当初の完全なシーケンスよりもよりコンパクトな動画コンテンツ表示である動画フレームの固定シーケンスのみを提供する。動画シーケンスを通して観察しコンテンツが変化するに従って、動画シーケンス

における関心度レベルも変化するので、大抵の場合、上記の表示はまだ不適切である。また、特定の動画コンテンツの関心度レベルは予測できない。少女が猫を可愛がっており、カメラが少女から猫にパンする動画シーケンスを考える。1人の視聴者は少女ではなく猫をより接近して見たいと思い、もう1人の視聴者は猫ではなく少女をより近くで見たいと思い、さらにもう1人の視聴者は両方を見たいと思うかも知れない。目標は、これらの視聴者の誰かが見る“興味の無いフレーム”の数を最小にすることである。

#### 【0046】

同じ動画インデックスシステムに対する種々の視聴要求を調和させ満足させるために、マルチ解像度動画ブラウザ（図2のブロック53）を提供してユーザが指定レベルの要約を選択して階層的要約を閲覧できるようにしている。これは、単なるインデックスシステムに代わるブラウザである。視聴者は、詳細の粗いレベルから始め、視聴者にとってより興味深いキーフレームシーケンス部分でマウスをクリックして詳細を拡張することができる。詳細1レベルより多い明細レベルが要求されて、視聴者が視聴者の選択したベースで閲覧できる。最も細密なレベルのキーフレームも検出できる。より粗いレベルにおいて、細密レベルにおける同様なキーフレームが集合し、各クラスは1枚の代表キーフレームにより代表される。

#### 【0047】

このクラスタリング問題を解決するために、よく知られているL i n d e e B u z o - G r a y (L B G) のアルゴリズム（又はL l o y dのアルゴリズム又はK平均アルゴリズム）の変更を提案する。同様な画像をクラスタさせることが望ましいことに留意する。ヒストグラムにより画像を表現し、類似の画像は類似のヒストグラムを有すると仮定する。各ヒストグラムを関連フレームの特徴ベクトルとして扱い、最も細密なレベルのNヒストグラムに代わる粗いレベルにおける代表的ヒストグラム（ $N/r$ ）を発見する。尚、Nは最も細密なレベルにおけるキーフレーム数である。パラメータ‘r’はコンパクション率（簡潔化率、若しくは、圧縮率）であり、ユーザによりプログラムに供給されるパラメータである。これからの議論において、各キーフレームをそのヒストグラムのベクトルと

して表現する。

#### 【0048】

例えば、連続な  $p$  個のベクトル（時間において連続）に代わる 1 個の代表的ベクトルを選び出すことが望ましいので、これは、正規のクラスタリング問題と異なっている。正規の LGB の場合は、1 つの代表ベクトルに量子化したベクトルに関する“連続性”の制限は何もない。正規の LBG 反復に近似の、下記の反復法は常に収斂する。この 3 ステップ反復法は、ここでは“対ごと”の LBG 又は PLBG と記述する。PLBG は LBG と同じ極小問題を有することに注意すべきである。幸運なことに、反復後の“クリーンアップステップ”をこれを迅速処理するために用いることができる。まず、ヒストグラムベクトル列を等間隔で区画することから始める。例えば、簡潔化率 3 の場合、各区画は、（最後の 1 つ又は 2 つの区画を除き）3 つのヒストグラムベクトルを含んでいる。次に、PLBG 法の次のステップに進む。

#### 【0049】

1. 各ベクトルセットに対して代表ベクトルとして重心（又は平均）ヒストグラムを指定する。
2. 最初の区画から開始し、各区画を、区画の何れかの側の 2 つの隣接セット（用語“対ごとの”はここに由来する）に対する合計基準  $l_2$  が最小になるように調節する。数学的に、 $H_{i-1}$  がセット  $(t_{i-1}, t_i)$  中のベクトルに対する代表ベクトルで、 $H_i$  がセット  $(t_i, t_{i+1})$  中のベクトルに対する代表ベクトルであれば、 $t_i$  を各セット中のベクトルの対応代表ベクトルまでの自乗距離の総計が最小になるように調節する。
3. 前記調節に続き、任意区画につき、 $t_i = t_{i+1}$  であれば、ベクトルの代表セットから  $H_i$  を除去する。 $t_{i-1} = t_i$  であれば、代表ベクトルセットから  $H_{i-1}$  を除去する。
4. ステップ 1 に戻る。

停止基準は、歪の減少量又は固定反復回数の何れかに基づく。前述の通り、10 回反復後停止する。各反復時に、歪（各セットの代表ベクトルとセット中の対応ベクトル間の基準  $l_2$ ）は減少する。かように、各反復における総歪は、減少

シーケンスを形成する。さらに、歪は、常にゼロより大きいかゼロに等しい。従って、シーケンスは、基本実解析による限界を有する。“反復に対する極小（従って固定点）が存在するか？”と云う問題は、純粋に学術的な問題であり、読者は、かような議論のための文献を参照することになる。除去ステップ（ステップ3）は、実際に、当初期待又は選択したより若干であるが少ないキーフレーム数となる結果を得た。

#### 【0050】

上記の方法において、停止後、代表ベクトルに最も近似したヒストグラムベクトルを有する第1クラスタ中のフレームを第1キーフレームとして選択する。後続クラスタに対するキーフレームも同様にして決定できる。“先行キーフレームとの最大差分判定基準”により後続クラスタ内でキーフレームを選択すると最良の結果が得られる。ここで、差分は動き測度で表現される。

#### 【0051】

上記反復の公式化において、最終区画が常にシーケンス中の最後のベクトルに固定されるので、最終セットが不適當に表現される可能性がある。同じことがショットの第1フレームに対しても言えるが、ここで報告する実験においては、そのような状況は観察されなかった。かように、この問題を解決するために、反復終了後にもう1つのステップを設ける。この最終ステージにおいて、代表ベクトルの最後にもう1つ代表ベクトルを付け加える必要があるかテストする。特に、最後のベクトルを新しい代表として加えることを検討する。最後のベクトルと先行代表ベクトルとの差が $\theta X$ （連続代表ベクトルの全ての対（ペア）間の差の平均）より小さければ、最後のベクトルの追加を認める。シミュレーション時 $\theta$ を0.75に選定する。 $\theta$ は0から1までの間で変化できる。

ベースライン法（Legendijk）は、3枚キーフレームを指定し、ベースライン法を動画シーケンスに適用すると、関心を引く特徴をもつシーンを見失う。その結果は、上記の方法を用いて生成した3枚キーフレームによる複数レベル階層の最もコンパクトな（最も粗い）レベルよりも劣る。さらに、複数レベルの要約を生成するために異なる数のキーフレームを得るには、ベースラインアルゴリズムを複数回用いるよりも、提案の階層的方法を用いる方がさらに有効であ

る。

#### 【0052】

##### ブロックヒストグラムの動き測度

ヒストグラムに基づく動き測度は全ての状況に適する訳ではない。例えば、黒い対象が白い背景に対して移動すると、ヒストグラムに基づく動き測度はその運動を記録できない。細かな動き、例えば、手のジェスチャ又は頭の動きを捕捉したい状況では、より良い動き測度を得る有利さを持っている。

#### 【0053】

ブロックヒストグラムはショットの検出のために提案された。しかしながら、ブロックヒストグラムはショットの検出には感度が高すぎて、多数の偽りの警告を発する。ブロックヒストグラムの背後にあるアイデアは、画像を若干数のブロック（通常4又は16）に分割し、動き測度を各ブロックに対する絶対値のヒストグラム差の合計をとって規定する。ブロックヒストグラムが単純な全体ヒストグラムに基づく方法では、捕捉できない動きに対してより鋭敏であることは容易に観察できる。ブロックヒストグラムを図4の3ステップ反復法に示すように、最も細密なレベルのキーフレームに対してのみ適用した。ブロックヒストグラムを用いる利点は、1つのフレームの代わりに動画フレーム当たり4又は16のヒストグラムを取り扱う必要があるので、計算及びメモリに関し、より集中的であることである。しかしながら、実験シーケンスにおいては、このブロックヒストグラム法は重要な性能改善をもたらさなかった。

#### 【0054】

##### 要約のための動き特性の使用

パン又はズームのような特に重要なケースはこれまで検討されていない。カメラパンの場合、インテリジェントブラウザは、（a）一回のパンで複数のフレームを検出し、（b）観察目的のために画像モザイクに変換するパンフレームに対するオプションを提供する。パンとズームのどちらの検出も動きベクトルの計算を含んでいるので、パンの検出と共にズームの検出は、大量の追加計算による過負荷を生じることなく実現できる。

#### 【0055】



シーケンス内の各フレームに対する動きベクトルを見つけ出すには、計算を要するので、支配的な又はグローバルな動きを持つフレームの全ての可能なシーケンスを先ず検出する予備スクリーニング（予備選抜）法を開発する。支配的な動きは（a）パン又は（b）ズーム又は（c）他の特別な編集効果により発生するので、検出したシーケンスをより詳細に調べ、パン又はズームの存在を決定する。

#### 【0056】

##### 支配的な動きのための予備スクリーニング

支配的な動きは、動画フレーム内の各画素が輝度の変化を経験することを意味する。輝度の変化は通常ズーム又はカメラの動きによって生じる。この変化は、動画フレームのエッジ画素において最も顕著である。この方法は、各画素を観察し、その画素がエッジ画素かどうかを決定し、エッジ画素であれば、現画素と先行フレーム中の同一位置の画素との間の差を検出する。エッジ画素における差の絶対値が閾値（ $PZ\_THRESH=15$ ）より大きければ、その画素は動きを有すると指定される。画素がエッジ画素かどうかを決定するために、その画素においてSobelエッジ検出演算子により得た数値を閾値（ $PZ\_THRESH=50$ ）と比較する。もし、 $PZ\_THRESH$ を減じると、偽の警報が寄せられることがある。 $PZ\_THRESH1$ を減じると、画素は強いエッジに属しておらず、動きが大きな輝度変化を生じないので、かような画素における有意の変化は無くなる。特別なフレームがパンフレームであるかを決定するために、エッジ画素の総数に対する、動きを有するものとして分類される画素数間の割合（パン率）に関する閾値を設ける（ $PZ\_THRESH2=0.9$ ）。

#### 【0057】

この割合が、パンの間、一貫して $PZ\_THRESH2$ を横断するのを確認するのに必要なもう1つのステップは、近傍にふくらませることである。言い換えれば、 $NEIGH \times NEIGH$ （ $NEIGH=5$ ）内の任意画素の輝度変化が $PZ\_THRESH$ より大きければ、エッジ画素は動きを有する。特定数より短いフレームシーケンスは拒絶される（ $TOO\_MANY\_FRAMES\_NOT\_PANZOOM=5$ ）。サブサンプリングは、計算負荷をさらに軽減するために

用いられる。

#### 【0058】

##### パン検出

このパン検出法は公知技法のバリエーションである。パンを検出するために、サブサンプリングした画像位置における動きベクトルを考察する（SPACING=24）。動きベクトルを決定するために使用する方法は、単純なブロックマッチング法である（BLKSIZE=7x7、SEARCHSIZE=24x24）。先行フレームのパンベクトルに基づく探索サイズを変化させる。先行のパンベクトルが（SEARCHSIZE/2）-2より小さければこの探索サイズを半分し、先回のパンベクトルが（SEARCHSIZE/2）-2より大きければ当初の（大きな）探索サイズを回復する。このバリエーションにより性能の低下が生じることはない。

#### 【0059】

パン検出の場合、許容範囲内のモード動きベクトルと（最も頻繁に生じる）平行な全ての動きベクトルを見つけ出すことが提案されている。かような動きベクトルの数が特定の閾値より大きければ、パンが検出される。しかしながら、1回のパンの場合、動きベクトルは並行しているだけでなく、同一に近い大きさを有している。従って、全並行動きベクトルを考察する代わりにモード動きベクトルの小さな近傍を調べる。モード動きベクトルの値に等しいベクトルが生じれば、恣意的な決定がなされる。モード動きベクトルの近傍サイズはVARN（=4）によって制御される。VARNの値が大きければ、モード動きベクトル周りの近傍は小さくなる（我々の場合のVARN=4は3x3の近傍を意味する）。PANRATIO（パン率）（=.5）は、動きベクトル総数に対する近傍内動きベクトル数間の割合に関する閾値を決定する。パンフレームシーケンス中の幾つかのフレームが閾値より低い場合でも、ホールが3より大きくない（TOO\_BIG\_A\_HOLE=3）であれば、パンの継続性が確保される。

#### 【0060】

##### ズーム検出

画像中の動きベクトル最も外側のリム、即ち、画像エッジの動きベクトルを調

べて、ズーム条件を検出すべきである。リムの直径方向で対向する位置における動きベクトルは反対方向を指示しなければならない。反対方向に向いている動きベクトルの動きベクトル総数に対する割合の閾値は ( $ZOOMRATIO = .7$ ) である。ズームの中心は画像の範囲内のどこに在ってもよいから、外側リム上の動きベクトルのみを使用する。かように外側リムの動きベクトルはズームの存在を示す最良のインジケータである。加えて、画像エッジには前景の動きがそれ程無い。

【0061】

#### 色処理

ここでは、先に開示した方法を色シーケンスに拡張する。2つの異なる実施形態を記述する。第1実施形態において、256ピンのYヒストグラムと2つの128ピンのUとVのヒストグラムより成る連結ヒストグラムを使用する。第2実施形態においては、256ピンのYヒストグラムを使用する。幾つかの実験シーケンスの場合、結果として重要な変化は何も観察されなかった。両方の場合とも、活動測度は式1によって規定される。しかしながら、色ヒストグラムにおいて若干のシーケンスを用いることは、2枚の動画フレーム間の変化を検出するのに決定的な役割を演じる。例えば、輝度(ルミナンス)は略同一であるが彩度値(クロマ)は変化する。

【0062】

#### 非圧縮動画入力方法の要約

図2は階層的要約と閲覧方法のブロック図である。ディゾルブ、フェードイン／フェードアウト、除去モジュールについては、参照のために引用した関連用途において説明し、ここではディゾルブ検出方法を開示する。このモジュールは、動画シーケンスから遷移フレームを除去してディゾルブを突発移行シーンに変換するために用いられる。最も細密なレベルのキーフレームを検出するブロックを主要ステップと共に図4に示す。図3は自動パン／ズーム自動処理モジュールの詳細を示している。自動的に、(a)パンの動きがあれば検出してモザイク(パノラマ)画像を構築し、(b)ズームシーケンスの最初と最後のフレームを検出し確認する。最も細密なレベルのキーフレームからパン／ズーム関連キーフレー

ムを除外して非パン／非ズームフレームだけを階層的要約キーフレーム生成プロセスに参加させる。この除去及び自動パン／ズーム処理は、オプションであり、必要に応じ、ユーザが動画クリップのある特定点において対話方式で実行可能にすることができる。このGUIにより、ユーザは、特定レベル要約の（階層的ブラウザにより生成された種々のレベルの内の）閲覧を開始することができる。例えば、モザイク画像とズーム要約と共に最も粗い要約を最初に表示することができる。次に、ユーザは、より細密な階層レベルにおけるフレームを対話方式で検索／再生することができる。ユーザは、ボタンをクリックして、現在見ているキーフレームの親－子の何れかにアクセスすることができる。親を選ぶと、現在のレベルのキーフレーム群が、親である1枚のキーフレームに代わる。子を選ぶと現在のキーフレームの子に該当する全てのキーフレームが見られる。図5は、親と子のキーフレームのコンセプトを示している。最も細密なレベルをさらに拡張し、即ち、最も細密なレベルにおける子では、指定のキーフレーム間の動画クリップを再生する。この時、動画が再生され、動画クリップのその部分に対応する音声と同時に再生される。動画クリップを再生するこの機能性はより粗いレベルの階層においても装備できる。

#### 【0063】

ここに記述した動画閲覧方法は、マルチメディア操作の有効なユーザインターフェースを提供するだけでなく、種々の用途に用いられる。第2世代の画像符号化システムを連想させると第2世代の動画符号化システムにおいて採用される可能性のある動画シーケンスの時間的特性に関する知見を提供する。例えば、MPEG-2ビットストリームを処理する設計の復号器はIBBP又はIBBBPフォーマットに適応する。しかしながら、キーフレームの階層を用いれば、時間的な動画ストリームの本質に、知的に、より重要に、計算効率よく適応し、少ない資源を用いて高い品質を供給する符号器を設計できる。圧縮の改善に動画フレームの階層を用いる方法に関する情報は文献で得られ、セグメンテーションアルゴリズムのマルチスケール特性をロスのない静止画像の圧縮を得るために活用されている。第2世代画像符号化システムと第2世代動画符号化システムの主たる違いは、前者は符号化機構の根本的な変更を必要とし、従って、大きな影響を

持ちえず、一方後者は現存の動画符号化基準のいずれにも組み込むことができる。

#### 【0064】

##### 計算性能

キーフレーム生成方法の計算性能は、本発明の方法を実施するために使用するコンピュータのハードディスクへのアクセス速度にかなり依存する。下記の議論において、“リアルタイム処理”は所定解像度において毎秒30フレームを処理する能力を意味する。300フレーム用1/4共通中間フォーマット(QCIF)のカラーシーケンス(176x144の解像度)の場合、SUN(登録商標)Ultra SPARC-2(登録商標)を用い、ヒストグラムの構築には11秒を要し、残りの処理には1秒未満を要することが判明した。かように、ヒストグラムの計算は、リアルタイムで達成できるとしても、リアルタイムで階層キーフレームの生成を容易に達成できなければならない。ヒストグラムの計算後の処理は実際のフレーム解像度とは関係なく、300フレームQCIFシーケンスを処理するに要した時間量は、各フレームのヒストグラムが事前に計算されていれば、1024x780の解像度でシーケンスを処理するに要する時間と同じであることも注記しておく。

現在のグローバルな動きの検出は、リアルタイムで実行できる。しかしながら、パン/ズーム検出に必要なブロックマッチングアルゴリズムを伴う厳しい計算負荷のために、パン/ズーム処理は実装したソフトウェアを用いてリアルタイムでは実行できない。

#### 【0065】

##### 圧縮動画入力

前述の論議において、非圧縮又は伸長ビットストリームのみを検討し実験に用いてきた。しかしながら、入手できる動画ストリームの殆どは、コンパクトに蓄積するために圧縮フォーマットされている。図4の方法を圧縮ビットストリームに拡張し、最少の復号を行いキーフレームを抽出できるようにする。圧縮動画を処理するブルートフォース方法で全体動画ストリームを簡単に伸長し、その後ここに記述する非圧縮動画用の技法を用いるのが良いと判断される。

## 【0066】

本項では、MPEG-2ビットストリームで利用できる、デジタル動画の階層要約と閲覧のバリエーションを記述する。全体のスキームは図7のフローチャートに要約して示す。ヒストグラムを計算する新規の方法を開示する。8 x 8ブロックのDC係数のヒストグラムを用いる。このプロセスは、入力ビット132により開始する。Iピクチャに対するヒストグラムの計算134は従って通常の技術を有する者にはよく知られた方法により直接的に行われる。予測ピクチャ（P及びBフレーム）に対するヒストグラムの計算は、参照フレームを完全に復号することなく、後述のようにして実行され、ヒストグラムの精度の向上と従ってキーフレームの選択136の精度向上がもたらされる。階層的キーフレームの選択136は、階層的要約のキーフレームの識別を、例えば、時間的な表示順序で決定し、この要約の情報を後述する復号器マネージャに供給する。DC T係数のヒストグラムが生成されると、階層的キーフレームの選択が、図4を参照し教示された様にして実行される。

## 【0067】

私の同時係属出願“動画シーケンス中のディゾルブ領域の検出（Detecting Dissolve Regions in Video Sequences）”に開示した機構のような動画中のディゾルブ領域検出機構を、ヒストグラムの計算とビットストリームインデックステーブル（BIT）の生成を行う図7の処理ブロック134に容易に組み込むことができる。即ち、ディゾルブ領域に含まれるフレームをBIT内にマークし、逐次キーフレーム選択プロセスにおいて無視する。或いは、ディゾルブ領域の内のフレームを偽のキーフレームにしてもよい。

## 【0068】

本方法は、ヒストグラムの計算134と同時に、ビットストリーム内のバイトオフセット位置、参照フレーム、DC Tブロック量子化に使用した量子化マトリックスのような各ピクチャに関する情報を含む、ビットストリームの記録を生成する。本発明において、“ビットストリームインデックステーブル”（BIT）と称するテーブルを生成する。BIT138のコンテンツとBITの生成方法に

については、後で詳述する。

B I T 1 3 8 の1つの目的は、ビットストリームの本質的なパラメータを捕捉する全ビットストリームを復号又はパーシング（構文解析）する必要なく要約を生成するためのキーフレームの復号を可能にすることである。構文解析は、システムに動画ストリームを復号するか否かにして、動画ストリーム中の各ビットを考察することを要求する。さらに、ユーザが、例えば、要約の表示又は2枚のキーフレーム間の動画の再生を望む際に効率的に閲覧を可能にするために、図8に示すように、要約の識別と原動画ビットストリームに加えB I T又はB I Tの縮小バージョンを装備する。ユーザに要約を提示し、若干の対話を含めた本方法の特別な実施態様について後述する。図8において、ビットストリームはB I Tと要約フレームの識別子とは物理的に異なる位置に置かれたメモリに常駐させることができる。例えば、ビットストリームをデータベースサーバに保管し、要約とB I Tをローカルマシーンに常駐させてもよい。代案として、3種のデータを全て、同一媒体、例えばD V Dディスク又は他の任意大容量蓄積媒体に常駐させることができる。B I Tをさらにコンパクトにする（又は削減する）方法については、“コンパクトな蓄積装置用の削減されたビットストリームインデックステーブルの生成”の章で説明する。B I Tを生成しB I Tサイズの“削減”を決定し、ダウンサイズB I Tに任意数の技法を使用できることは評価すべきである。一つの例をここにあげる。B I Tを形成保管せず、全ビットストリームを構文解析して復号を要するキーフレームを毎回復号することもまた可能である。

#### 【0069】

図7を再び参照する。階層的要約140の生成期間中、復号器マネージャ142がB I Tに含まれている情報を利用しキーフレームを選択的に復号してM P E G - 2復号器144に送り、復号次第、階層要約140を形成する。復号器マネージャ142は提示段階中に、ユーザがキーフレーム間の動画を再生して動画を閲覧することを望む際に、同様なタスクを実行する。復号器マネージャの作動原理（例えば、コンピュータプログラムにより実装される）について次に説明する。

#### 【0070】

本発明は、MPEG-2圧縮動画を蓄積して記録する動画カメラ内に実装できる。かような場合、要約情報とBITは動画ストリームを蓄積する蓄積システムに蓄積できる。或いは、要約情報とBITを、明確な方式で動画ストリームと連結された任意のメモリ位置に蓄積してもよい。階層的要約自体は、キーフレーム又はキーフレームをサブサンプリングしたバージョンを含んでおり、迅速にアドレスできる蓄積システムに保管できる。オンカメラ式ユーザインターフェイスは、カメラ、テープ又は任意の他の蓄積媒体に階層的要約に基づき保管された動画コンテンツの認定のために装備される。

#### 【0071】

別案として、ビットストリームをカメラから要約プロセスを実行するコンピュータにダウンロードしてもよい。この場合、要約はテープ又は動画データを保持する任意の他の蓄積媒体又はビットストリームとの明確なリンクで連結された他のメモリにコピーする。例えば、圧縮されたMPEGストリームを直接記録するカメラが現在入手可能であり（例えば、日立MP-EG1A型カメラ）、ビットストリームをパーソナルコンピュータ（PC）にダウンロードできる。本発明のシステムは、PCプラットフォームでかようなビットストリームを処理するために使用できる。

#### 【0072】

MPEG-2ビットストリームを効率よく処理し階層的動画要約を作成するためには、以下の問題に取り組み解決しなければならない。

1. MPEG-2ビットストリームを最小限復号しながら、キーフレームの階層を生成する。
2. 全てのフレームを復号することなくMPEG-2ビットストリームから選択したキーフレームを復号する手順を確立する。
3. 2枚の所与のキーフレーム間の一連のフレームを復号する戦略を開発する。

この方法は、ヒストグラムレベルで働く。MPEG-2ビットストリームを最小限復号しながらフレーム毎に色ヒストグラムを計算する方法を開示する。

#### 【0073】

ヒストグラムの計算と後続処理は、各次元における8分の1のサブサンプリン



グから大きな影響を受けない。1ステップ進めて、 $8 \times 8$ ブロックの平均である、 $8 \times 8$ ブロックDCTのDC成分のみ使用して計算したヒストグラムは実際の目的には十分であった。動き補償画像の場合、計算量を削減するために、性能の低下が無視できる程度である近似動き補償を用いることができるとして提案されている。MPEG規格に用いられているブロックマッチングのスキームに従い、 $16 \times 16$ マクロブロックの動きベクトルを、実行中の動き補償予測の参照フレームの $16 \times 16$ マクロブロックの最大限4つのマクロブロックと重ね合わせる。同様に、 $16 \times 16$ マクロブロック中の各 $8 \times 8$ のサブブロックを最大限4つの他の $8 \times 8$ サブブロックと重ね合わせる。かように、各 $8 \times 8$ サブブロックを、重なる各 $8 \times 8$ サブブロック中の値の重み付けた平均値で近似させ得ることが示唆された。個別ブロックに割り当てられた重み値は、重なり領域に比例させることができた。図9を参照し、 $8 \times 8$ サブブロックの平均値は、次式で計算される。

【0074】

【式5】

$$((a)(b)(m_1) + (8-a)(b)(m_3) + (a)(8-b)(m_2) + (8-a)(8-b)(m_4) + \Delta_{DCT}) / 64 \quad (5)$$

【0075】

ここで、 $\Delta_{DCT}$  は、ブロックに対する残差DCTのDC成分の8倍である（MPEG-2規格に用いられるブロックに対する残差DCTのDC成分はブロックの残差誤りの平均値の $1/8$ であるので、8の係数が入る）。ヒストグラムは、式5中に見られるように、画像内の各 $8 \times 8$ ブロックの平均でヒストグラムベクトルを更新することにより得られる。ヒストグラムを得る上記の方法は、性能の低下をもたらす問題を有している。本発明の方法の改良の1つは、MPEGビットストリームに対するヒストグラムの計算を取り扱うより良い方法を提案することにある。

MPEGビットストリームは、特定のフレームを復号しようとする前にビットストリームの他の部分からの復号情報を必要とする複雑な符号化戦略を内蔵している。上首尾の動画閲覧方法も最少の時間量で特定の動画フレームを復号する間

題を検討する必要がある。

【0076】

MPEGビットストリームからのヒストグラムの計算

MPEGビットストリームの復号は、次の2つの計算が集中するステップを含んでいる。

1.  $8 \times 8$  ブロックの逆DCTを得る。
2. MPEG-2の場合、 $16 \times 16$  マクロブロックによる動き補償。ブロックは小さいか又は偶数／奇数フィールドのみを有している。

先に、 $8 \times 8$  ブロックをその平均値で置き換えても画像のヒストグラムに大きな影響を及ぼさないことを指摘した。この実装例において、各 $8 \times 8$  ブロックを $8X$  (DCT係数のDC値) で置換する。逆DCT計算のための公式から、これはブロックの平均値を生じ、補償関連量子化誤差の範囲内で正確である。

【0077】

次のステップを理解するために、MPEGビットストリームに採用されている符号化戦略についてその概略を説明する。典型的なMPEGビットストリームは3種類のフレームを有している。

- I (フレーム内符号化フレーム)、
- B (双方向予測フレーム) 及び
- P (予測フレーム)。

I フレームはDCTデータのみを含んでいる (動き補償は行わない)。かように、ヒストグラムの計算にDCT係数のDC値を用いるとIフレームの最少限復号の問題は完全に解決できる。BフレームとPフレームは先に復号した参照フレームから現フレームを予測するためにブロック動きベクトルを用いる追加ステップを含む。先に復号した利用可能なフレーム自身は部分的にのみ復号されている。かように、BフレームとPフレームの復号に用いられる戦略は、十分に検討しなければならない。以下の検討において、ケース (a) は既に存在し文献において共通に用いられている動き補償スキームに関し、ケース (b) はここで開示する新しい動き補償スキームに関する。

【0078】

動き補償を単純化するために、最も知られている方法で、前節で記述したスキームを用い、即ち、各8 x 8サブブロックを、それが重なった8 x 8サブブロックの重みを付けた平均値で置き換える。2つのシナリオを検討する。ケース（a）は8 x 8サブブロックを部分的に復号した参照フレーム中の重なったブロックの重みを付けた平均値で置き換える。ケース（b）は、8 x 8サブブロックを、部分的に復号した参照フレームからの正確な画素で置き換える。ケース（a）において、動き補償予測フレーム中の全8 x 8ブロックが単一値を持つことが判る。ケース（b）においては、8 x 8ブロックは潜在的に多くの異なる値を持ち得る（即ち、ブロック内の画素は多くの異なる値を取り得る）。これをさらに説明するために、ケース（a）とケース（b）を通して8 x 8ブロックの例を検討する。図10はこの例を示している。図10において、予測ブロックはIフレームから得る、即ち、8 x 8ブロックは予測フレーム中の関連する単一値を持つものと仮定する。ケース（a）は、ただ1つの値 $\mu$ を持つ現フレーム中の1つの8 x 8ブロックを生じる。ケース（b）は潜在的に4つの異なる値を有する現フレーム内の1つの8 x 8ブロックを生じる。

#### 【0079】

これは、フレーム内符号化参照フレーム（又はIフレーム）に続く第1の幾つかの動き補償フレーム（PまたはBフレーム）に大きな差を生じない。事実、ヒストグラムの計算は平均化及びサブサンプリングから大きな影響を受けないので、2つの手順はヒストグラムの計算に対しては同等に有効であるとみなされる。しかしながら、ケース（a）は計算量が少なくメモリの消費も少ないのでより好ましい。これは、ケース（a）のように動き補償された任意所与のフレーム（I又はP又はB）においては、各8 x 8ブロックにはただ1つの値が得られるからである。かように、各次元において8分の1の容量を要するだけで済む。即ち、全フレームを蓄積する場合と比較して潜在的可能性としては、64（8 x 8）分の1の容量のメモリ装置ですむことになる。しかしながら、以下に説明するように、ケース（a）は過度の性能低下をもたらすので、使用可能な代案ではない。

#### 【0080】

連続動き補償フレーム数、Iフレームを介在させない、すなわち連続動き補償

フレーム数が増加すると、ケース（a）とケース（b）間の差が増大する。図3に戻って、1枚の動き補償済みフレームから予測を、例えば、PフレームからBフレーム又はPフレームからもう1つのPフレームに予測を行う際に、何が生じるかを考察する。ケース（a）の場合、重み付け平均化処理を予測ブロックが重なる4つのブロックに対して実施すると、各ブロックは単一の値を持っており、 $8 \times 8$ の予測ブロック全体に対し単一の値で終了する。ケース（b）の場合は、予測フレームの各ブロックは潜在的に4つ（又はそれ以上）の異なる値を有するので、現予測ブロックは多数の異なる値を持ち得る。いま、動き補償済みフレームからのこの一連の予測としてケース（a）とケース（b）間に出現する重要な変化キーとなる差分が大きくなることに注目する。動き補償フレームのシーケンスが十分に長いと、ケース（a）は、後述するようにフレーム間の動きが十分であれば、フレーム全体に対し単一の値を生じると期待される。これは、しかしながら、ケース（b）には生じない。

#### 【0081】

この現象をさらに詳細に説明するために、Iフレーム内の各 $8 \times 8$ ブロックを平均で置換し、原画像のより小さいバージョンを生成することを考える。今、ケース（a）において実施したように動き補償は、この小画像を1個の $2 \times 2$ 平均化フィルタで反復再帰処理することを必要とする。基本的なフーリエ解析で、平均化フィルタの反復用途によりエッジ効果を無視すれば制限内の均質な画像を生じることを容易に示すことができる。

#### 【0082】

上記の観察が真実であることが実際に観察された。典型的なMPEG-2圧縮シーケンスの場合、2枚のIフレーム間の距離は15である。これにより、動き補償をケース（a）に従い実行すると非常に顕著な性能低下が生じることが発見された。ケース（a）の動き補償スキームはヒストグラムに周期的に強い変動を生じさせ、偽のキーフレームを検出する結果をもたらす。かような理由で、ケース（b）を実装に使用した。最少復号法を用いヒストグラムを計算することにより、1つのQCIFシーケンスに対するヒストグラム計算速度を半分に短縮されるが、利点は解像度が高まることの方が大きかった。現在、512次元のヒスト

グラムのベクトルを使用しており、256ピンのグレイスケール（Y成分）ヒストグラムと128ピンのU成分ヒストグラムと128ピンのV成分ヒストグラムにより形成される。上記の説明は、彩度（クロマ）フォーマットとは無関係に、Y、U及びVのフレーム構成成分に対して個々に適用できる。

#### 【0083】

##### MPEG-2ビットストリームからの特別フレームの抽出

ここに開示する実施形態において、MPEG-2ビットストリームからの特定フレームの抽出は2ステップの手順で行われる。ヒストグラムの計算と同時に実行される第1ステップにおいて、MPEG-2ビットストリームからランダムに選んだフレームを迅速に復号するために必要な情報を含む“ビットストリームインデックステーブル”を生成する。キーフレームの階層が生成される、即ち、階層的要約に入るキーフレームの識別が規定されると、階層の最も細密なレベルのキーフレームだけが復号を要求され、階層の粗いレベルのフレームは最も細密なレベルのフレームのサブセットとなる。キーフレーム抽出手順における第2ステップは、後述する復号器マネージャにより、第1ステップで生成された“ビットストリームインデックステーブル”を使用して実行される。

#### 【0084】

上記2ステップ手順の利点は、ビットストリームの関連部分を直接復号する場合に比し、関心を引くフレームまでの全ビットストリームを見るのに要する時間を削減できることである。フレーム番号1350から1400までのMPEG-2ビットストリームをビットストリームインデックステーブルを用いずに復号するためには、フレーム番号1350までの全ビットストリームを、完全に復号しないにしても、構文解析することが必要である。これにはかなりの時間を要する。ビットストリームインデックステーブルが使用できれば、ビットストリームの関連部分に直接進むことができ、絶対最少量の構文解析と復号のみが要求される。

#### 【0085】

MPEG-2ビットストリームからランダムに選択したフレーム（ここでは現フレームと記す）を復号するためには下記の情報が必要である。

1. 過去において（時間的に）最も近いシーケンスのヘッダ（そのバイトオフセット）。
2. 現フレームのビットストリームへのバイトオフセット。
3. 過去において（そのバイトオフセット）最も近い量子化マトリックスのリセット（もしあれば）。
4. 現フレームがBフレームであれば、現フレームに対応する参照フレーム（I-P/I-I/P-P）（それらのフレームのバイトオフセット）。
5. 現フレームがPフレームであれば、（Pフレームの参照フレームである）最も近いIフレーム（そのフレームのバイトオフセット）。

## 【0086】

B/Pフレームの場合は、参照フレームを正確に復号するために、参照フレーム以外の幾つかの他の参照フレームを復号する必要があることを指摘しておく。

前記2つのステップ、即ち、（1）ビットストリームインデックステーブルを生成するステップと（2）後述の復号器マネージャによるビットストリームインデックステーブルを用いたフレームの抽出するステップ間で情報交換が容易に行えるように、上記情報を保持するための共通のデータ構成を開発した。Cコードの下記のセグメントを使用しビットストリームインデックステーブルの形成に用いる異なるフラグを実態化した。但し、これは一例に過ぎず、ビットストリームインデックステーブルは任意数の異なるシンタックス形式をもち得ることは理解されよう。

## 【0087】

【表1】

```
enum IndexFileState {
    K_SEQUENCE_HEADER=0,
    K_PICTURE_IFRAME,
    K_PICTURE_BFRAME,
    K_PICTURE_PFRAME,
    K_QUANT_MATRIX_EXTENSION,
    K_END_OF_DATA,
    K_OFFSET
};
```

## 【0088】

K\_OFFSETフラグは、上記の他のフラグから区別するために任意バイトオフセットに加える。15のバイトオフセットは、我々の表記では15+K\_OFFSET(=21)と翻訳する。K\_END\_OF\_DATAは、異なるイベント(例えば、シーケンスヘッダとIフレーム又はIフレームとBフレーム等)間の区切りである。生成されたビットストリームインデックステーブルがどのように出現するかを理解するために、下記イベントのシーケンスの符号化が要求されていると仮定する。

## 【0089】

1. シーケンスヘッダが0バイトで開始
2. Iピクチャが150バイト
3. Pピクチャが3000バイト
4. Bピクチャが4200バイト
5. Bピクチャが5300バイト
6. 量子化マトリックスのリセットが5400バイト
7. Pピクチャが6200バイト

このシーケンスを下記の表現に変換する。

## 【0090】

【表2】

```

K_END_OF_DATA K_SEQUENCE_HEADER K_OFFSET+0 K_END_OF_DATA
K_PICTURE_IFRAME K_OFFSET+150 K_END_OF_DATA
K_PICTURE_PFRAME K_OFFSET+3000 K_END_OF_DATA
K_PICTURE_BFRAME
K_OFFSET+4200 K_END_OF_DATA K_PICTURE_BFRAME K_OFFSET+5300
K_QUANT_MATRIX_EXTENSION K_OFFSET+5400 K_END_OF_DATA
K_PICTURE_PFRAME K_OFFSET+6200 K_END_OF_DATA

```

## 【0091】

これは前記のCデータ構成を用いて下記のバイト表現を生成する。

## 【0092】

【表3】

5 0 6 5 1 156 5 2 3006 5 3 4206 5 3 5306 4 5406 5 2 6206 5

【0093】

上記バイトを用いた表現中のスペースは、復号器がビットストリームを構文解析するために必要である。K\_\_END\_\_OF\_\_DATAフラグは、厳密に言えば冗長である。しかしながら、このフラグは、発生した偽のデータを除去するために使用でき（ビットストリーム中のエラーによって）、このアルゴリズムをエラーに対し強くする。このフラグは偽データを除去するための“同期信号”として作用し、例えば、（ビットストリーム中のエラーによる）バイトオフセット値を伴わないシーケンスヘッダを破棄する。

【0094】

上記の生成された“ビットストリームインデックステーブル”を使用する復号器マネージャは下記のように機能する。

1. 最後に復号されたフレームの番号（l d f）を-1に初期化する。
2. 復号される各フレーム（f t d）に対し：

復号を開始するフレーム（s d f）を見つけ出す。

f t dがIフレームであれば、s d f = f t dとする。

f t dがPフレームであれば、s d fをf t dより前の最も近いIフレームにセットする。

f t dがBフレームであれば、s d fをf t dに対応するアンカーフレームの両方より前の最も近いIフレームにセットする。

上記s d fがl d f + 1より小さければ、s d f = l d f + 1に設定する。s d f < l d f + 1であれば、要求されたフレームの幾つかはすでに復号されている。

i = l d f + 1 ~ s d fに対して、

最も近いシーケンスヘッダ（r s h）を見つけ出す。

q m rがr s hより大きい場合は、最も近い量子化マトリックスのリセット（q m r）を見つけ出す（q m rが存在すれば）。

r s hとq m rをビットストリーム中に出現する順序で復号する。



全てのI及びPフレームを、s d fから開始し、f t d - 1に到るまで順次復号する。

f t dを復号する。

1連のフレームを復号するために、復号器マネージャの上記手順に従って、第1フレーム(f t d)を復号する。残りのフレームは、1連のフレームの最後まで順次復号する。

#### 【0095】

特別な場合に必要に応じ、所要のフィールドピクチャを処理する。偶数/奇数フィールドヒストグラムを使用することができ、どちらも最初に復号できる。他のフィールドは、Bピクチャの場合復号できず、P及びIピクチャの場合は、最少符号化戦略で復号できる。1フィールドのみ復号する場合は、ヒストグラムを2分の1に縮尺しなければならない。余分に復号したフィールドはP/Iフレーム用ヒストグラムの計算には使用せず、この場合、全てのフレームはヒストグラムに1フィールド供給しているので、ヒストグラムを縮尺しなくてよい。フィールドをフレームから区別し適切なステップをとるために、MPEG-2ビットストリームは、ピクチャヘッダとピクチャ符号の拡張から下記の2つの情報を提供する。

1. (ピクチャヘッダ中の) 時間的な基準が現在復号中のフレーム番号を提供する。この時間基準は各グループのピクチャヘッダの先頭にリセットされる。
2. (ピクチャ符号の拡張中の) ピクチャ構成が、トップ/ボトムフィールド情報を提供する。

#### 【0096】

#### コンパクトな蓄積装置用の削減されたビットストリームインデックステーブルの生成

実装の観点から重要な問題は、ディスクスペースを節減するためにビットストリームインデックステーブルをコンパクトに表現することである。MPEG動画が占拠する大スペースと比べると、ビットストリームインデックステーブルは各フレーム当たり約8-10バイトのスペースをとるだけなので、これは、一見、重要な問題ではないよう見えるかもしれない。オーバーヘッドは下記のステップ

を取ることににより低減できる。

1. 絶対バイトオフセットよりも増分バイトオフセットを用いる。これにより大きなシーケンスの場合かなりの量の節減になる。
2. UnixプラットフォームのgzipまたはPC上のpkzipのような圧縮アルゴリズムを用いる。
3. K\_\_END\_\_OF\_\_DATAフラグを除去する。
4. ビットストリームを最少の復号及び構文解析してキーフレームを復号するために必要な最少量の情報を蓄積するように、ビットストリームインデックステーブルのデータ量を削減する。

#### 【0097】

最後の項は、キーフレームの位置にのみアクセスでき、キーフレームの位置をユーザが後で変更することを許さないことに注意すべきである。(1)と(4)は完全には両立しないことも注記しておく。増分バイトオフセットを用いる場合、(4)の削減アルゴリズムは、フレームの変更順序を反映するようにバイトオフセットを変更すべきである。現在の実装において、削減されたビットストリームインデックステーブルを後述のように生成する。復号器マネージャがこの削減バージョンのテーブルを使用する。この削減ビットストリームテーブルの使用は、得られるべき速度と使用可能な蓄積スペースの量の関数として決定される。蓄積装置のスペースが使用可能であれば、ビットストリームテーブルを削減する必要はない。完全なテーブルを納める余地があれば、検索及び処理も迅速に行える。

#### 【0098】

##### 削減されたビットストリームインデックステーブル

ビットストリームインデックステーブル(BIT)と同じ構成を用いるが、同時に削減ビットストリームインデックステーブル(PBIT)用の異なる組織シンタックスを採用している。各キーフレームは、BIT内の1単位として表現されている各動画フレームに対向し、(K\_\_END\_\_OF\_\_DATAのフラグで挟まれた)1単位として表現される。全ビットストリームを構文解析及び復号することなく特定の(現)キーフレームを復号するためには、下記の情報が必要であ

る。

【0099】

1. 現キーフレームのタイプI/B/P
2. 現キーフレームに対応する開始復号フレーム(sdf) (バイトオフセット)。復号器マネージャのBITバージョンにおいて仮定したように、復号器は全てのキーフレームを復号するためには順次作動しないので、 $(sdf < ldf + 1)$ であれば $ldf + 1$ でない) バイトオフセット表示の実際のsdfは保管する必要がある。復号器が順次作動し、 $sdf < ldf + 1$ であれば、復号は $ldf + 1$ から開始することが必要である。 $ldf + 1$ が先行キーフレームに続く動画フレームであり、そのオフセットは入手できるので、この情報は既に入手可能である。
3. 現キーフレームのバイトオフセット
4. 最も近いシーケンスヘッダのオフセット。全ての有効なMPEG-2ビットストリームはシーケンスヘッダを有しているので、K\_SEQUENCE\_HEADERフラグは必要ない。
5. 量子化マトリックスのリセットがあった場合は、MPEG-2ビットストリームには量子化マトリックスリセットがないので、K\_QUANT\_MATRIX\_EXTENSIONフラグを付けて蓄積する必要がある。シーケンスヘッダの出現で量子化マトリックスがリセットされるので、そのリセットがそのシーケンスヘッダ以前に発生していればその量子化マトリックスのリセットを蓄積する必要はない。

【0100】

復号器マネージャは、sdf情報と現キーフレームのタイプとそのバイトオフセットを次のように用いる。所望の現キーフレームがI又はPタイプであれば、復号器マネージャはsdfから復号を開始し、ビットストリームを構文解析し、I及びPフレームのみを探す。かようなI及びPフレームは、所望の現フレームに到達するまで復号され、所望の現フレームも復号される。この技法において、復号器マネージャは、フレームがBフレームあるかをチェックせず、I及びPフレームのヘッダのみを探索する。所望の現キーフレームがBタイプであれば、復

号器マネージャはs d fから始まる各フレームを考慮し、全てのI及びPフレームを復号し、各Bフレームで停止してそのフレームが所望のキーフレームであるかをチェックする。

#### 【0101】

そのフレームがBタイプであれば、Bフレームを構文解析する必要を無くすために、最も近い復号参照（I又はP）フレームに近似させたい。Bフレームが53500バイトのオフセット値より始まる復号すべきキーフレームであり、43000バイトから読む量子化マトリックスを必要とし、39000バイト位置で読むシーケンスヘッダを必要としていると仮定する。復号を開始するフレームは45000バイトより始まる。このデータは次のように符号化されている。

#### 【0102】

【表4】

```
K_END_OF_DATA K_PICTURE_BFRAME K_OFFSET+45000
K_OFFSET+53500 K_OFFSET+39000
K_QUANT_MATRIX_EXTENSION K_OFFSET+43000 K_END_OF_DATA
```

#### 【0103】

復号器マネージャが実行する構文解析量が増大することにより計算時間が増えてコストが掛かるためにP B I Tをさらにスリムにできることを明らかにしておかねばならない。従って、復号器マネージャの複雑化とP B I Tのサイズとの間にトレードオフがある。用途の要求条件により適当なバランスをとることも可能である。例えば、P B I Tはs d fと、s d fと現キーフレーム間の全てのI及びPフレームのバイトオフセットと、現フレームのバイトオフセットを蓄積し、復号器マネージャがビットストリームを構文解析する必要を無くす。但し、その結果、P B I Tのサイズは増大する。他の極端な場合は、s d fとキーフレームのバイトオフセットのみを蓄積する。最もコンパクトなP B I Tが得られるが、その代わり、復号器マネージャは、開始フレームからキーフレーム位置までのビットストリームを構文解析し、IとPのフレームを復号しなければならない。自動パン／ズーム検出／抽出機能を組み込むためには、全フレームビットストリームを復号する必要がある。

## 【0104】

デジタル動画シーケンスのキーフレームを閲覧するシステムを開示してきた。入力動画ストリームは、通常のデジタル動画であるか、或いは、DCTに基づく圧縮ストリームである。本発明の好適な実施態様と幾つかの変更例を開示してきたが、特許請求範囲に規定した本発明の範囲から逸脱することなく、さらなる変更及び修正を加えることができることを理解すべきである。

## 【図面の簡単な説明】

## 【図1】

図1は、3つのレベルに対する動画の要約の階層構造を示す図である。

## 【図2】

図2は、本発明による方法の第1実施形態のブロック図である。

## 【図3】

図3は、本発明の自動パン／ズーム処理モジュールのブロック図である。

## 【図4】

図4は、本発明の最も細密なレベルのキーフレームの選択アルゴリズムを示すブロック図である。

## 【図5】

図5は、本発明の階層的要約のブロック図である。

## 【図6】

図6は、キーフレーム ( $k_j$ ) および対応ショットセグメント ( $t_{j-1}$ ,  $t_j$ ) の累積動き測定値 ( $C(x)$ ) と分布を示す図である。

## 【図7】

図7は、MPEG-2による圧縮入力動画に使用される本発明の第2実施形態の一部を示すブロック図である。

## 【図8】

図8は、階層的要約中のキーフレームの復号に用いられるデータの示す図である。

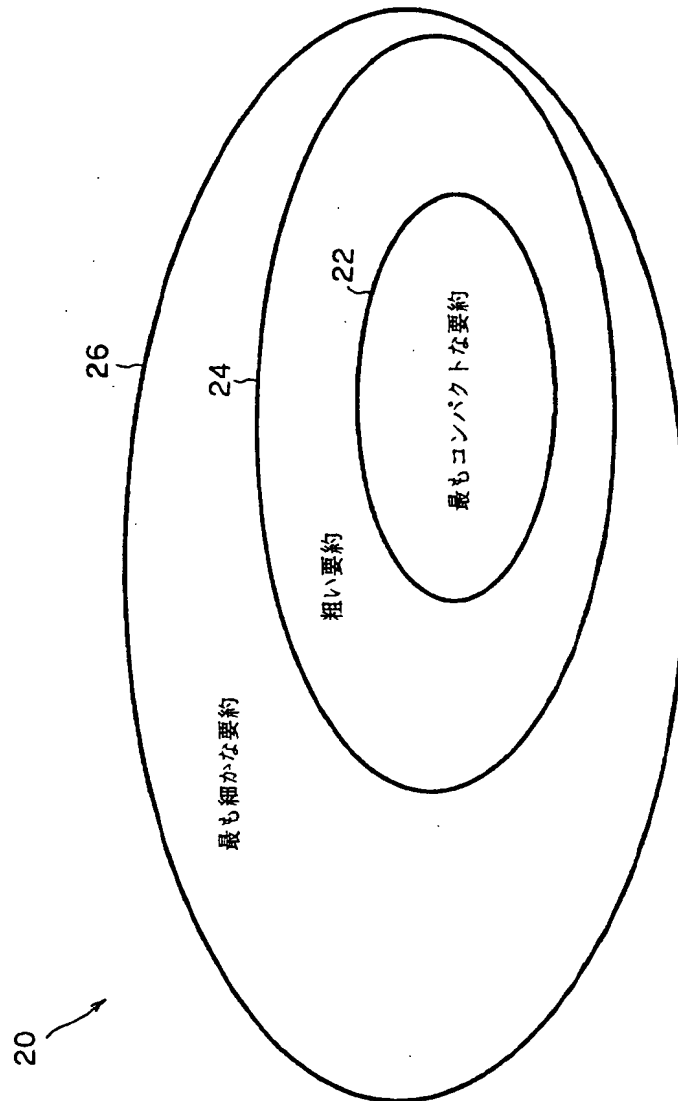
## 【図9】

図9は、動き補償のグラフである。

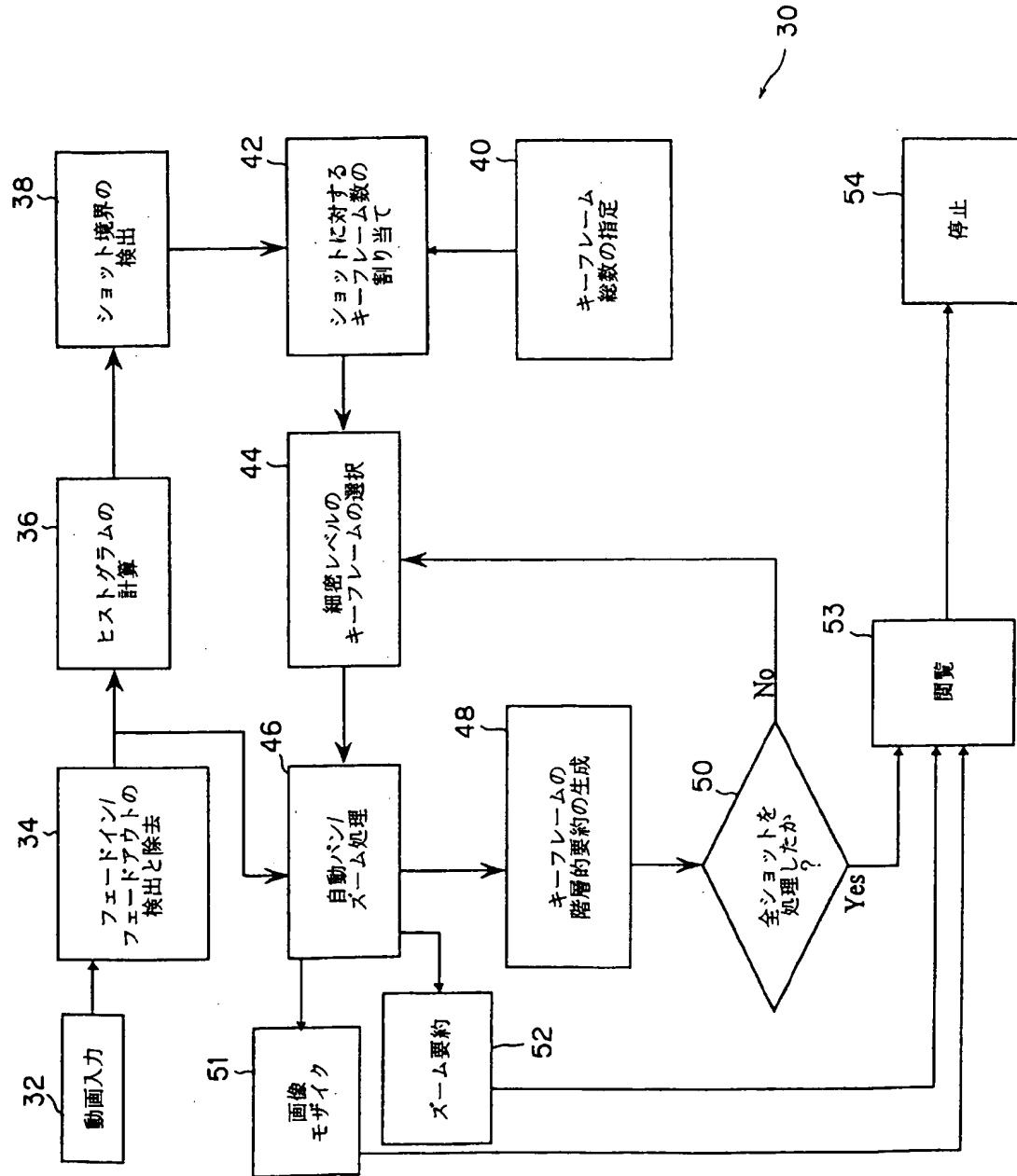
【図10】

図10は、ケース（a）及びケース（b）を定義するために用いる動き補償アルゴリズム間の差分を示す図である。

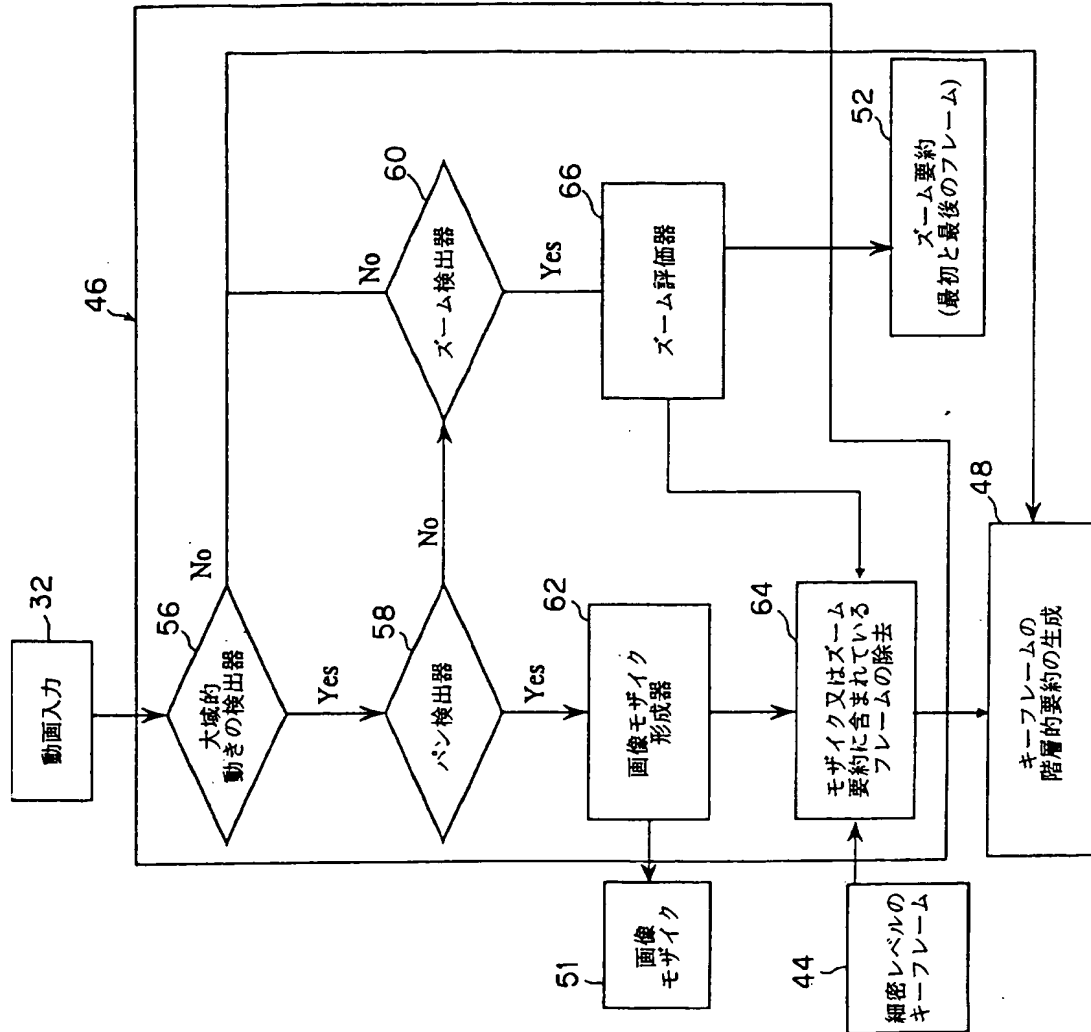
【図1】



【図2】

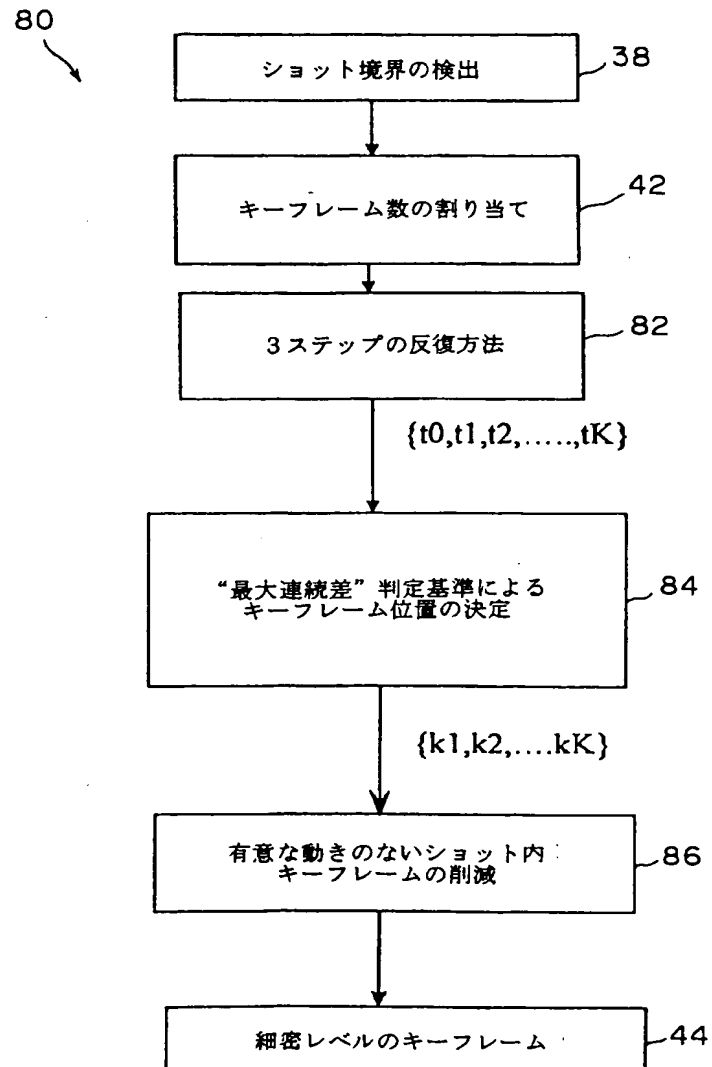


【図3】

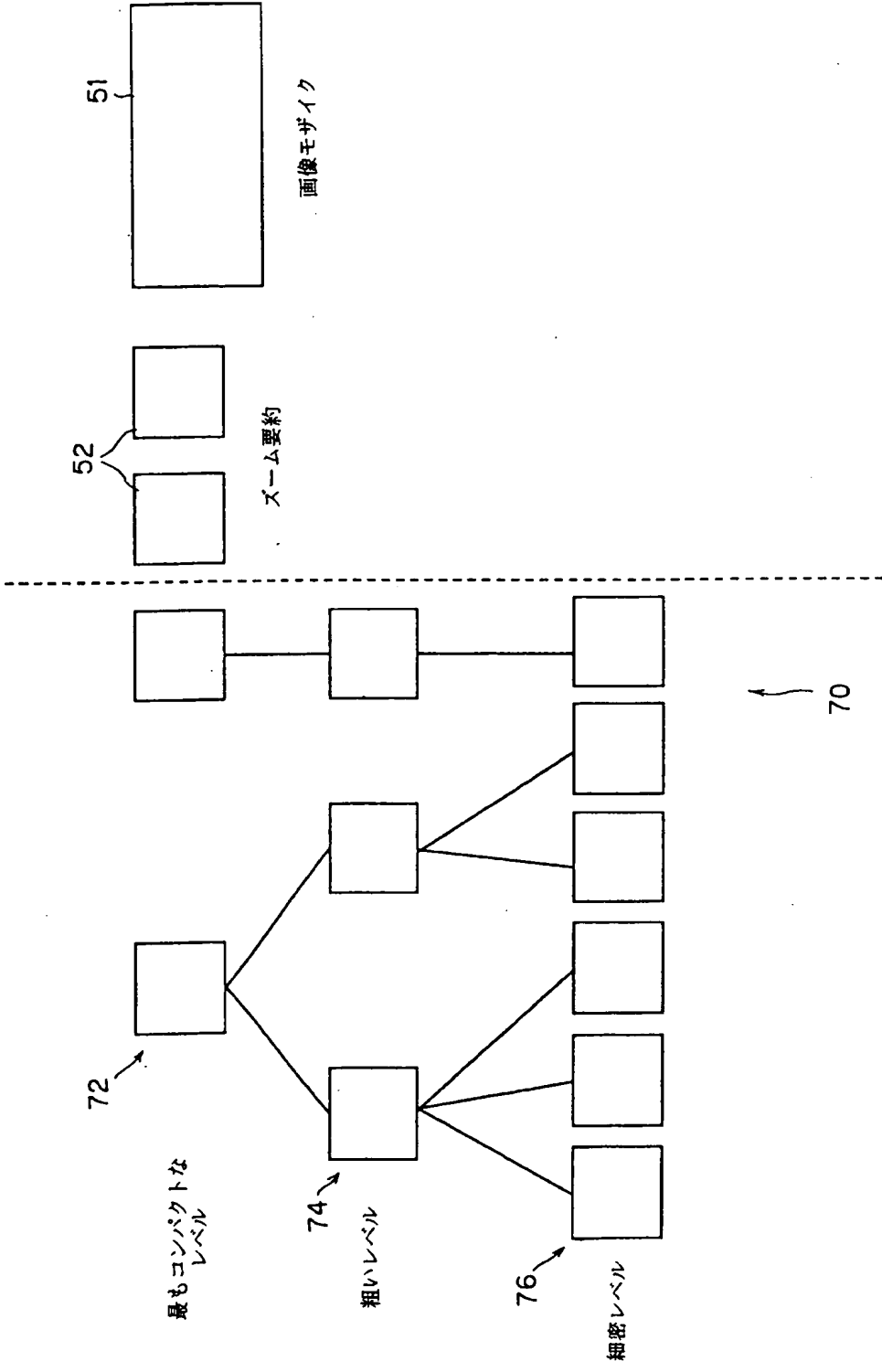




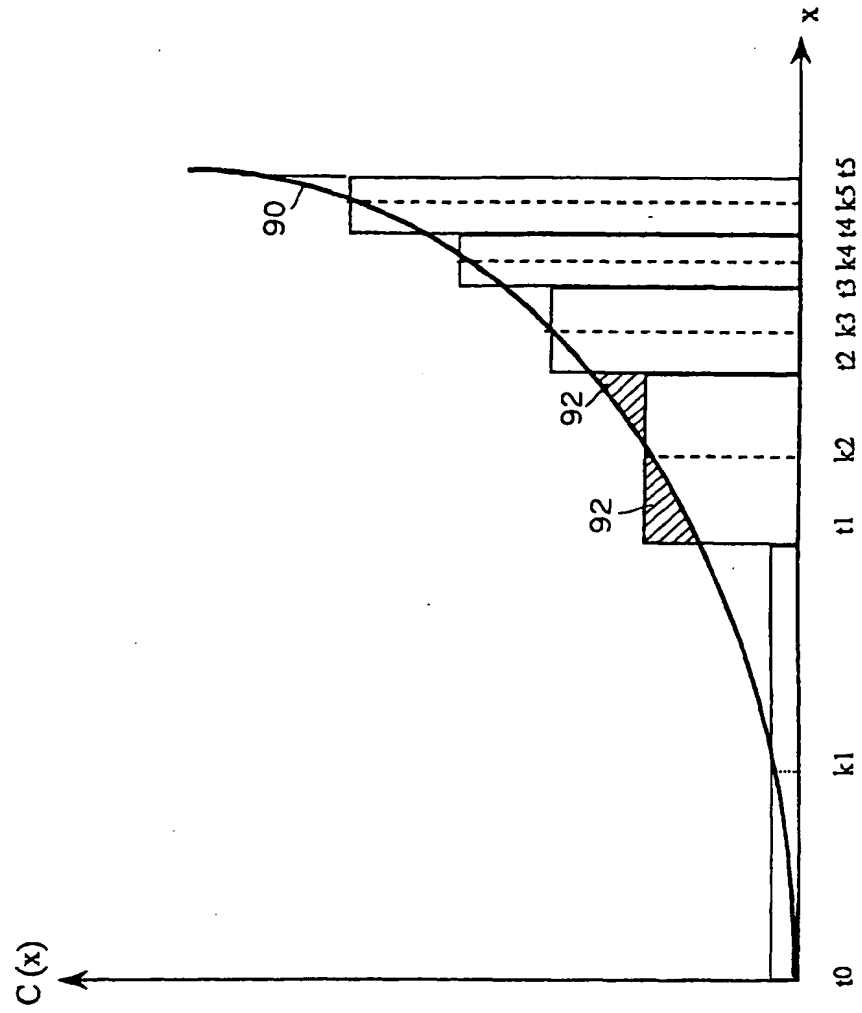
【図4】



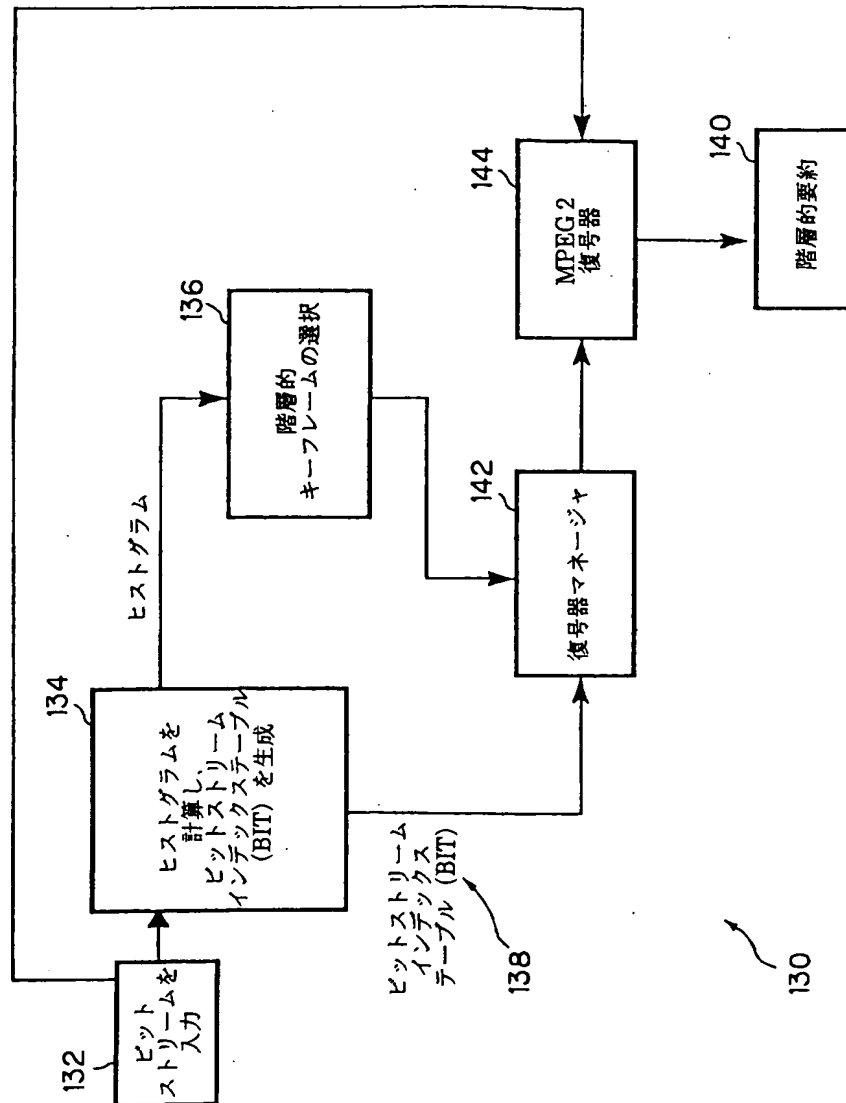
【図5】



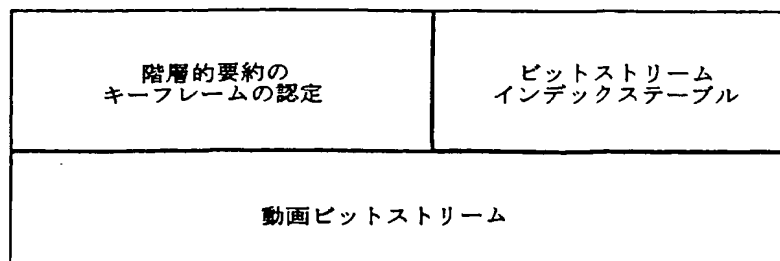
【図6】



【図7】

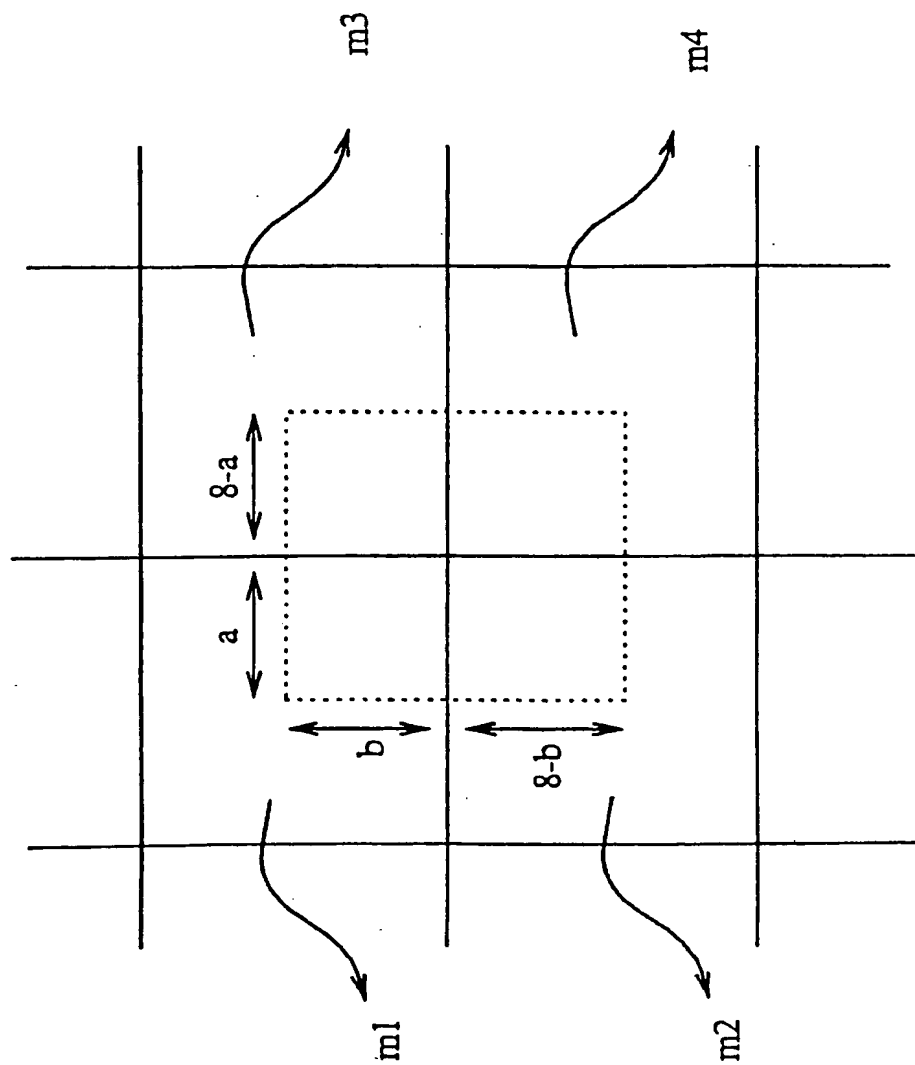


【図8】

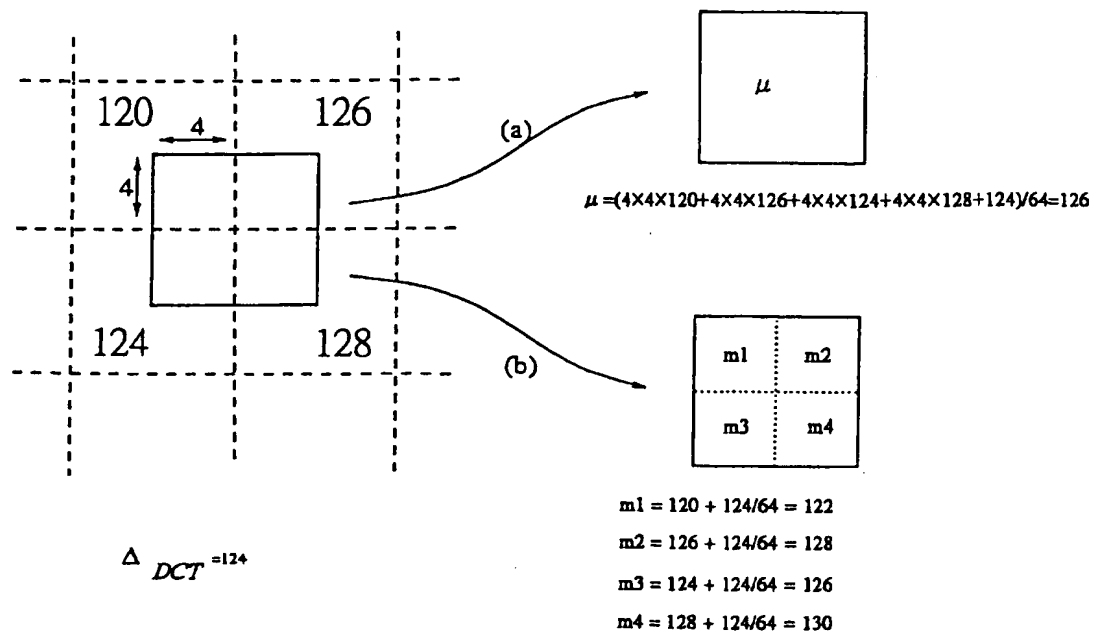


特表 2001-527304

【図9】



【図10】



【手続補正書】特許協力条約第34条補正の翻訳文提出書

【提出日】平成11年11月10日（1999. 11. 10）

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】請求項1

【補正方法】変更

【補正内容】

【請求項1】 デジタル動画シーケンスに対するデジタル動画信号を入力するステップと、

前記デジタル動画シーケンスに対するヒストグラムを計算するステップと、

前記デジタル動画シーケンス内のショット境界を検出するステップと、

各ショット内に割り当てられたキーフレーム数を決定するステップと、

各キーフレームの動きのある局面を位置決めするステップであって、最も大きい測度において先行キーフレームの動き測度と異なる動き測度を持つフレームを前記動画シーケンスのキーフレーム指定するステップと、

ショットから選択されたキーフレームを削減するステップであって、選択されたキーフレームが先行フレームと次のキーフレームの動き測度から予め定められた測度だけ逸脱しない動き測度をもつキーフレームであるステップと、

階層的キーフレームの要約を生成するステップを含んでなることを特徴とする階層的デジタル動画要約及び閲覧方法。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】請求項3

【補正方法】変更

【補正内容】

【請求項3】 前記生成した後に、前記階層的フレームの要約を用いるキーフレームを閲覧することを含むことを特徴とする請求項1に記載の階層的デジタル動画要約及び閲覧方法。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】請求項4

【補正方法】変更

【補正内容】

【請求項4】 前記入力した後に、ディゾルブイベントを検出し、除去することを含むことを特徴とする請求項1に記載の階層的デジタル動画要約及び閲覧方法。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】請求項20

【補正方法】変更

【補正内容】

【請求項20】 前記デジタル動画シーケンスは、圧縮されたデジタル動画シーケンスであり、前記入力するステップは、ビットストリームのインデックステーブルを生成することをさらに含み、前記計算するステップは、前記圧縮されたデジタル動画シーケンスを部分的にのみ復号することを特徴とする請求項1に記載の階層的デジタル動画要約及び閲覧方法



## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

International Application No. PCT/JP 98/05715		
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 6 G06F17/30 G11B27/10		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 6 G06F G11B G06T		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No.	
P,X	US 5 708 767 A (YEO BOON-LOCK ET AL) 13 January 1998 see abstract; figures 1,4 see column 6, line 35-37 ---	1-3
A	US 5 532 833 A (HONG JUNG-KOOK ET AL) 2 July 1996 see abstract; figures 1,3,4 ---	1,3
A	EP 0 555 874 A (INTEL CORP) 18 August 1993 see page 3, line 8-12; claims 1,2 --- -/-	2,4
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		
<input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (see specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search 26 Apr11 1999	Date of mailing of the international search report 04/05/1999	
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx: 31 651 epo nl Fax: (+31-70) 340-3016	Authorized officer Jonsson, P.O.	

Form PCT/ISA/210 (second sheet) (July 1992)

## INTERNATIONAL SEARCH REPORT

Int'l. Application No.  
PCT/JP 98/05715

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SAWHNEY H S ET AL: "COMPACT REPRESENTATIONS OF VIDEOS THROUGH DOMINANT AND MULTIPLE MOTION ESTIMATION" IEEE TRANS. ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 18, no. 8, August 1996, pages 814-830, XP000632862 see abstract see paragraph 2.4 see paragraph 5.2	5-8
A	NAKAJIMA Y: "A VIDEO BROWSING USING FAST SCENE CUT DETECTION FOR AN EFFICIENT NETWORKED VIDEO DATABASE ACCESS" IEICE TRANS. ON INF. AND SYSTEMS, vol. E77-D, no. 12, 1 December 1994, pages 1355-1364, XP000497103 see abstract	1
A	see paragraph 3; figure 6	9
A	see paragraph 4; figure 7	20
A	CHEN H -Y ET AL: "A MULTI-LAYER VIDEO BROWSING SYSTEM" IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, vol. 41, no. 3, August 1995, pages 842-849, XP000539544 see abstract; figures 5-7	1,9

## INTERNATIONAL SEARCH REPORT

Information on patent family members

Int'l. Application No.

PCT/JP 98/05715

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5708767 A	13-01-1998	NONE	
US 5532833 A	02-07-1996	NONE	
EP 0555874 A	18-08-1993	US 5283645 A	01-02-1994
		JP 6070232 A	11-03-1994

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.